(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
31 May 2001 (31.05.2001)

PCT

(10) International Publication Number
WO 01/38577 A2

(51) International Patent Classification[7]: C12Q 1/68

(21) International Application Number: PCT/US00/31922

(22) International Filing Date:
21 November 2000 (21.11.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/448,480 24 November 1999 (24.11.1999) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 09/448,480 (CON)
Filed on 24 November 1999 (24.11.1999)

(71) Applicant *(for all designated States except US)*: **THE JOHNS HOPKINS UNIVERSITY** [US/US]; Suite 906, 111 Market Place, Baltimore, MD 21202 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **VELCULESCU, Victor, E.** [US/US]; Apt. 903, 1 E. Chase St., Baltimore, MD 21202 (US). **VOGELSTEIN, Bert** [US/US]; 3700 Breton Way, Baltimore, MD 21208 (US). **KINZLER, Kenneth, W.** [US/US]; 1403 Halkirk Way, BelAir, MD 21015 (US).

(74) Agents: **HEMMENDINGER, Lisa, M.** et al.; Banner & Witcoff, Ltd., 11th floor, 1001 G Street, N.W., Washington, DC 20001-4597 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/38577 A2

(54) Title: HUMAN TRANSCRIPTOMES

(57) Abstract: Global gene expression patterns have been characterized in normal and cancerous human cells using serial analysis of gene expression (SAGE). Cancer cell-specific, cell-type specific, and ubiquitously expressed genes have been identified. This information can be used to provide combination of cell type- and cancer-specific gene probes, as well as methods of using these probes to identify particular cell types, screen for useful drugs, reduce cancer-specific gene expression, standardize gene expression, and restore function to a diseased cell or tissue.

HUMAN TRANSCRIPTOMES

This invention was made with government support under CA57345, CA62924, and CA43460 awarded by the National Institutes of Health. The government has certain rights in the invention.

BACKGROUND OF THE INVENTION

The characteristics of an organism are largely determined by the genes expressed within its cells and tissues. These expressed genes can be represented by transcriptomes that convey the identity and expression level of each expressed gene in a defined population of cells (**1, 2**). Although the entire sequence of the human genome will be elucidated in the near future (**3**), little is known about the many transcriptomes present in the human organism. Basic questions regarding the set of genes expressed in a given cell type, the distribution of expressed genes, and how these compare to genes expressed in other cell types, have remained largely unanswered.

General properties of gene expression patterns in eukaryotic cells were determined many years ago by RNA-cDNA reassociation kinetics (**4**), but these studies did not provide much information about the identities of the expressed genes within each expression class. Technological constraints have limited other analyses of gene expression to one or few genes at a time (**5-9**) or were non-quantitative (**10, 11**). Serial analysis of gene expression (SAGE) (**12**), one of several recently developed gene expression methods, has permitted the quantitative analysis of transcriptomes in the yeast *Saccharomyces cereviseae* (**1, 13**). This effort identified the expression of known and previously unrecognized genes in *S.*

cereviseae (1, 14) and demonstrated that genome-wide expression analyses were practicable in eukaryotes.

Thus, there is a need in the art for the identification of transcriptomes which represent gene expression in particular cell types or under particular physiological conditions in eukaryotes, particularly in humans.

SUMMARY OF THE INVENTION

It is an object of the present invention to provide such transcriptomes, individual polynucleotides, and methods of using the polynucleotides to identify particular cell types, screen for useful drugs, reduce cancer-specific gene expression, standardize gene expression, and restore function to a diseased cell or tissue. These and other objects of the invention are provided by one or more of the embodiments described below.

One embodiment of the invention is a method of identifying a cell as either a colon epithelial cell, a brain cell, a keratinocyte, a breast epithelial cell, a lung epithelial cell, a melanocyte, a prostate cell, or a kidney epithelial cell. Expression in a test cell of a gene product of at least one gene is determined. The at least one gene comprises a sequence selected from at least one of the following groups:

(a) the sequences shown in SEQ ID NOS:2, 5-18, 20-84, and 85;

(b) the sequences shown in SEQ ID NOS:87-96, 98, 100-103, 105, 107-110, 112-129, 131-150, and 151;

(c) the sequences shown in SEQ ID NOS:152-154 and 155;

(d) the sequences shown in SEQ ID NOS:156-159 and 160;

(e) the sequences shown in SEQ ID NOS:161-166 and 167;

(f) the sequences shown in SEQ ID NOS:168, 170, 172-177, 179-188, 190-207, and 208;

(g) the sequences shown in SEQ ID NOS:209 and 210; and

(h) the sequences shown in SEQ ID NOS:211-224 and 225.

Expression of a gene product of at least one gene comprising a sequence shown in (a) identifies the test cell as a colon epithelial cell. Expression of a gene product of at least one gene comprising a sequence shown in (b) identifies the test cell as a brain cell. Expression

of a gene product of at least one gene comprising a sequence shown in (c) identifies the test cell as a keratinocyte. Expression of a gene product of at least one gene comprising a sequence shown in (d) identifies the test cell as a breast epithelial cell. Expression of a gene product of at least one gene comprising a sequence shown in (e) identifies the test cell as a lung epithelial cell. Expression of a gene product of at least one gene comprising a sequence shown in (f) identifies the test cell as a melanocyte. Expression of a gene product of at least one gene comprising a sequence shown in (g) identifies the test cell as a prostate cell. Expression of a gene product of at least one gene comprising a sequence shown in (h) identifies the test cell as a kidney epithelial cell.

Another embodiment of the invention is an isolated polynucleotide comprising a sequence selected from the group consisting of SEQ ID NOS:2, 5, 6, 8, 10, 12, 13, 15, 17, 18, 21, 24-26, 28, 30, 31, 34-36, 38, 40, 47-51, 53-57, 59-62, 65-69, 71-76, 78, 80-84, 98, 103, 113, 115, 122, 129, 132, 134, 135, 140, 144, 149, 150, 153-168, 174-176, 182, 185, 186, 188, 190, 200, 201, 205-213, 216-224, 237, 239, 257, 263, 485, 487, 495, 499, 514, 586, 686, 751, 835, 844, 878, 910, 925, 932, 951, 1000, 1005, 1070, 1122, 1130, 1170, 1173, 1187, 1189, 1200, 1213, 1220, 1237, 1257, 1264, 1273, 1293, 1300, 1320, 1367, 1371, 1401, 1403, 1404, 1406, 1418, and 1419.

Still another embodiment of the invention is a solid support comprising at least one polynucleotide. The polynucleotide comprises a sequence selected from at least one of the following groups:

(a) the sequences shown in SEQ ID NOS:2, 5, 6, 8, 10, 12, 13, 15, 17, 18, 21, 24-26, 28, 30, 31, 34-36, 38, 40, 47-51, 53-57, 59-62, 65-69, 71-76, 78, 80-83, and 84;

(b) the sequences shown in SEQ ID NOS:98, 103, 113, 115, 122, 129, 132, 134, 135, 140, 144, 149, and 150;

(c) the sequences shown in SEQ ID NOS:153-154 and 155;

(d) the sequences shown in SEQ ID NOS:156-157 and 160;

(e) the sequences shown in SEQ ID NOS:161-166 and 167;

(f) the sequences shown in SEQ ID NOS:168, 174-176, 182, 185, 186, 188, 190, 200, 201, 205-207 and 208;

(g) the sequences shown in SEQ ID NOS:209 and 210;

(h) the sequences shown in SEQ ID NOS:211-213, 216-223, and 224;

(i) the sequences shown in SEQ ID NOS:237, 239, 257, and 263; or

(j) the sequences shown in SEQ ID NOS:485, 487, 495, 499, 514, 586, 686, 751, 835, 844, 878, 910, 925, 932, 951, 1000, 1005, 1070, 1122, 1130, 1170, 1173, 1187, 1189, 1200, 1213, 1220, 1237, 1257, 1264, 1273, 1293, 1300, 1320, 1367, 1371, 1401, 1403, 1404, 1406, 1418, and 1419.

Even another embodiment of the invention is a method of identifying a test cell as a cancer cell. Expression in a test cell of a gene product of at least one gene is determined. The at least one gene comprises a sequence selected from the group consisting of SEQ ID NOS:228, 230-257, 259-260, and 262-265. An increase in expression of at least two-fold relative to expression of the at least one gene in a normal cell identifies the test cell as a cancer cell.

Yet another embodiment of the invention is a method of reducing expression of a cancer-specific gene in a human cell. A reagent which specifically binds to an expression product of a cancer-specific gene is administered to the cell. The cancer-specific gene comprises a sequence selected from the group consisting of SEQ ID NOS:228, 230-257, 259-260, and 262-265. Expression of the cancer-specific gene is thereby reduced relative to expression of the cancer-specific gene in the absence of the reagent.

Even another embodiment of the invention is a method for comparing expression of a gene in a test sample to expression of a gene in a standard sample. A first ratio and a second ratio are determined. The first ratio is an amount of an expression product of a test gene in a test sample to an amount of an expression product of at least one gene comprising a sequence selected from the group consisting of SEQ ID NOS:266-375, 377-652, 654-796, and 798-1448 in the test sample. The second ratio is an amount of an expression product of the test gene in a standard sample to an amount of an expression product of the at least one gene in the standard sample. The first and second ratios are compared. A difference between the first and second ratios indicates a difference in the amount of the expression product of the test gene in the test sample.

Still another embodiment of the invention is a method of screening candidate anti-cancer drugs. A cancer cell is contacted with a test compound. Expression of a gene

product of at least one gene in the cancer cell is measured. The at least one gene comprises a sequence selected from the group consisting of SEQ ID NOS:228, 230-257, 259, 260, 262-263, and 265. A decrease in expression of the gene product in the presence of a test compound relative to expression of the gene product in the absence of the test compound identifies the test compound as a potential anti-cancer drug.

Still another embodiment of the invention is a method of screening test compounds for the ability to increase an organ or cell function. A selected from the group consisting of a colon epithelial cell, a brain cell, a keratinocyte, a breast epithelial cell, a lung epithelial cell, a melanocyte, a prostate cell, and a kidney cell is contacted with a test compound. Expression in the cell of a gene product of at least one gene is measured. The gene comprises a sequence selected from at least one of the following groups:

(a) the sequences shown in SEQ ID NOS:2, 5-18, 20-84, and 85;

(b) the sequences shown in SEQ ID NOS:87-96, 98, 100-103, 105, 107-110, 112-129, 131-150, and 151;

(c) the sequences shown in SEQ ID NOS:152-154 and 155;

(d) the sequences shown in SEQ ID NOS:156-159 and 160;

(e) the sequences shown in SEQ ID NOS:161-166 and 167;

(f) the sequences shown in SEQ ID NOS:168, 170, 172-177, 179-188, 190-207 and 208;

(g) the sequences shown in SEQ ID NOS:209 and 210; and

(h) the sequences shown in SEQ ID NOS:211-224 and 225.

An increase in expression of a gene product of at least one gene comprising a sequence shown in (a) identifies the test compound as a potential drug for increasing a function of a colon cell. An increase in expression of a gene product of at least one gene comprising a sequence shown in (b) identifies the test compound as a potential drug for increasing a function of a brain cell. An increase in expression of a gene product of at least one gene comprising a sequence shown in (c) identifies the test compound as a potential drug for increasing a function of a skin cell. An increase in expression of a gene product of at least one gene comprising a sequence shown in (d) identifies the test compound as a potential drug for increasing a function of a breast cell. An increase in expression of a gene product

of at least one gene comprising a sequence shown in (e) identifies the test compound as a potential drug for increasing a function of a lung cell. An increase in expression of a gene product of at least one gene comprising a sequence shown in (f) identifies the test compound as a potential drug for increasing a function of a melanocyte. An increase in expression of a gene product of at least one gene comprising a sequence shown in (g) identifies the test compound as a potential drug for increasing a function of a prostate cell. An increase in expression of a gene product of at least one gene comprising a sequence shown in (h) identifies the test compound as a potential drug for increasing a function of a kidney cell.

Yet another embodiment of the invention is a method to restore function to a diseased tissue. A gene is delivered to a diseased cell selected from the group consisting of a colon epithelial cell, a brain cell, a keratinocyte, a breast epithelial cell, a lung epithelial cell, a melanocyte, a prostate cell, and a kidney cell. The gene comprises a nucleotide sequence selected from at least one of the following groups:

(a) the sequences shown in SEQ ID NOS:2, 5-18, 20-84, and 85;

(b) the sequences shown in SEQ ID NOS:87-96, 98, 100-103, 105, 107-110, 112-129, 131-150, and 151;

(c) the sequences shown in SEQ ID NOS:152-154 and 155;

(d) the sequences shown in SEQ ID NOS:156-159 and 160;

(e) the sequences shown in SEQ ID NOS:161-166 and 167;

(f) the sequences shown in SEQ ID NOS:168, 170, 172-177, 179-188, 190-207, and 208;

(g) the sequences shown in SEQ ID NOS:209 and 210; and

(h) the sequences shown in SEQ ID NOS:211-224 and 225.

Expression of the gene in the diseased cell is less than expression of the gene in a corresponding cell which is normal. If the diseased cell is a colon epithelial cell, then the nucleotide sequence is selected from (a). If the diseased cell is a brain cell, then the nucleotide sequence is selected from (b). If the diseased cell is a keratinocyte, then the nucleotide sequence is selected from (c). If the diseased cell is a breast epithelial cell, then the nucleotide sequence is selected from (d). If the diseased cell is a lung epithelial cell,

then the nucleotide sequence is selected from (e). If the diseased cell is a melanocyte, then the nucleotide sequence is selected from (f). If the diseased cell is a prostate cell, then the nucleotide sequence is selected from (g). If the diseased cell is a kidney cell, then the nucleotide sequence is selected from (h).

Thus, the invention provides transcriptomes, polynucleotides, and methods of identifying particular cell types, reducing cancer-specific gene expression, identifying cancer cells, standardizing gene expression, screening test compounds for the ability to increase an organ or a cell function, and restoring function to a diseased tissue.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1. Sampling of gene expression in colon cancer cells. Analysis of transcripts at increasing increments of transcript tags indicates that the fraction of new transcripts identified approaches 0 at approximately 650,000 total tags.

FIG. 2. Colon cancer cell Rot curve.

FIGS. 3A-3C. Gene expression in different tissues. **FIG. 3A**. Fold reduction or induction of unique transcripts for each of the comparisons analyzed. The source of the transcripts included in each comparison are displayed in FIG. 3C. The relative expression of each transcript was determined by dividing the number of transcript tags in each comparison in the order displayed in FIG. 3C. To avoid division by 0, we used a tag value of 1 for any tag that was not detectable in one of the samples. We then rounded these ratios to the nearest integer; their distribution is plotted on the X axis. The number of transcripts displaying each ratio is plotted on the Y axis. Each comparison is represented by a specific color (see below or FIG. 3C). **FIG. 3B**. Expression of transcripts for each comparison, where values on X and Y axes represent the observed transcript tag abundances in each of the two compared sets. Light Blue symbols: DLD1 in different physiologic conditions; Yellow symbols: DLD1 cells (X axis) versus HCT116 cells (Y axis); Red symbols: colon cancer cells (X axis) versus normal brain (Y axis); and Dark Blue symbols: colon cancer cells (X axis) versus hemangiopericytoma (Y axis). **FIG. 3C**. Fraction of transcripts with dramatically altered expression. For each comparison, Expression Change denotes the number of transcripts induced or reduced 10 fold, and (%) denotes the number of altered

transcripts divided by the number of unique transcripts in each case. Differences between expression changes were evaluated using the chi squared test, where the expected expression changes were assumed to be the average expression change for any two comparisons.

TABLE LEGENDS

Table 1. Table of tissues and transcript tags analyzed. "Tissues" represents the source of the RNA analyzed, "Libraries" indicates the number of SAGE libraries analyzed, "Total Transcripts" is the total number of transcripts analyzed from each tissue, and "Unique Transcripts" denotes the number of unique transcripts observed in each tissue.

Table 2. Table of transcript abundance. "Copies/cell" denotes the category of expression level analyzed in transcript copies per cell, "Unique Transcripts" represents the number of unique transcripts observed and those matching GenBank genes or ESTs, and "Mass fraction mRNA" represents the fraction of mRNA molecules contained in each expression category.

Table 3. Table showing tissue-specific transcripts. The number in parentheses adjacent to the tissue type indicates the percent of transcripts exclusively expressed in a given tissue at 10 copies per cell. "Transcript tag" denotes the 10 bp tag adjacent to 4 bp *Nla*III anchoring enzyme site, "Copies/cell" denotes the transcript copies per cell expressed, and "UniGene Description" provides a functional description of each matching UniGene cluster (from UniGene Build No. 67). As UniGene cluster numbers change over time, the most recent cluster assignment for each tag can be obtained individually at http://www.ncbi.nlm.nih.gov/SAGE/SAGEtag.cgi (Lal *et al.*, "A public database for gene expression in human cancers," *Cancer Research*, in press) or for the entire table at http://www.sagenet.org/transcriptome.

Table 4. Table showing ubiquitously expressed genes. "Copies/cell" denotes the average expression level of each transcript from all tissues examined, "Range" represents the range in expression for each transcript tag among all tissues analyzed in copies per cell, and "Range/Avg" is the ratio of the range to the average expression level and provides a measure of uniformity of expression. Other table columns are the same as in Table 5. The

entire table of uniformly expressed transcripts also is available at http://www.sagenet.org/transcriptome.

Table 5. Table showing transcripts uniformly elevated in human cancers. Transcripts expressed at 3 copies/cell whose expression is at least 2-fold higher in each cancer compared to its corresponding normal tissue. CC, colon cancer; BC, brain cancer; BrC, breast cancer; LC, lung cancer; M, melanoma; NC, normal colon epithelium; NB, normal brain; NBr, normal breast epithelium; NL, normal lung epithelium; NM, normal melanocytes. "Avg T/N" is the average ratio of expression in tumor tissue divided by normal tissue (for the purpose of obtaining this ratio, expression values of 0 are converted to 0.5). Other table columns are the same as in Table 5.

Table 6. Table showing transcripts expressed in colon cancer cells at a level of at least 500 copies per cell.

Table 7. Table showing transcripts expressed at a level of at least 500 copies per cell.

DETAILED DESCRIPTION OF THE INVENTION

It is a discovery of the present invention that particular sets of expressed genes ("transcriptomes") are expressed only in cancer cells; expression of these genes can be used, *inter alia*, to identify a test cell as cancerous and to screen for anti-cancer drugs. These cancer-specific genes can also provide targets for therapeutic intervention.

It is another discovery of the invention that other transcriptomes are differentially associated with distinct cell types; expression of genes of these transcriptomes can therefore be used to identify a test cell as belonging to one of these distinct cell types.

It is yet another discovery of the invention that genes of another transcriptome are expressed ubiquitously; expression of genes of this transcriptome can be used to standardize expression of other genes in a variety of gene expression assays.

To identify the transcriptomes described herein we used the SAGE method, as described in Velculescu *et al.* (1) and Velculescu *et al.* (12), to analyze gene expression in a variety of different human cell and tissue types. The SAGE method is also described in U.S. Patents 5,866,330 and 5,695,937. A total of 84 SAGE libraries were generated from

19 tissues (Table 1). Diseased tissues included cancers of the colon, pancreas, breast, lung, and brain, as well as melanoma, hemangiopericytoma, and polycystic kidney disease. Normal tissues included epithelia of the colon, breast, lung, and kidney, melanocytes, chondrocytes, monocytes, cardiomyocytes, keratinocytes, and cells of prostate and brain white matter and astrocytes.

A total of 3,496,829 transcript tags were analyzed and found to represent 134,135 unique transcripts after correcting for sequencing errors (transcript data available at http://www.sagenet.org./transcriptome). Expression levels for these transcripts ranged from 0.3 to a high of 9,417 transcript copies per cell in lung epithelium. Comparison against the GenBank and UniGene collections of characterized genes and expressed sequence tags (ESTs) revealed that 6,900 transcript tags matched known genes, while 65,735 matched ESTs. The remaining 61,500 transcript tags (46%) had no matches to existing databases and corresponded to previously uncharacterized or partially sequenced transcripts.

Each of the genes or transcripts whose expression can be measured in the methods of the invention comprises a unique sequence of at least 10 contiguous nucleotides (the "SAGE tag"). Genes which are differentially expressed in colon, lung, kidney, and breast epithelial cells, brain cells, prostate cells, keratinocytes, or melanocytes are shown in Table 3. Ubiquitously expressed genes are shown in Table 4. Transcripts which are expressed only in cancer tissues, *e.g.*, colon cancer, breast cancer, brain cancer, liver cancer, and melanoma, are shown in Table 5.

This information provides heretofore unavailable picture of human transcriptomes. These results, like the human genome sequence, provide basic information integral to future experimentation in normal and disease states. Because SAGE analyses provide absolute expression levels, future SAGE data can be directly integrated with those described here to provide progressively deeper insights into gene expression patterns. Eventually, a relatively complete description of the transcripts expressed in diverse cell types and in various physiologic states can be obtained.

Isolated polynucleotides

The invention provides isolated polynucleotides comprising either

deoxyribonucleotides or ribonucleotides. Isolated DNA polynucleotides according to the invention contain less than a whole chromosome and can be either genomic DNA or DNA which lacks introns, such as cDNA. Isolated DNA polynucleotides can comprise a gene or a coding sequence of a gene comprising a sequence as shown in SEQ ID NOS:1-1563, such as polynucleotides which comprise a sequence selected from the group consisting of SEQ ID NOS:2, 5, 6, 8, 10, 12, 13, 15, 17, 18, 21, 24-26, 28, 30, 31, 34-36, 38, 40, 47-51, 53-57, 59-62, 65-69, 71-76, 78, 80-84, 98, 103, 113, 115, 122, 129, 132, 134, 135, 140, 144, 149, 150, 153-168, 174-176, 182, 185, 186, 188, 190, 200, 201, 205-213, 216-224, 237, 239, 257, 263, 485, 487, 495, 499, 514, 586, 686, 751, 835, 844, 878, 910, 925, 932, 951, 1000, 1005, 1070, 1122, 1130, 1170, 1173, 1187, 1189, 1200, 1213, 1220, 1237, 1257, 1264, 1273, 1293, 1300, 1320, 1367, 1371, 1401, 1403, 1404, 1406, 1418, and 1419.

Any technique for obtaining a polynucleotide can be used to obtain isolated polynucleotides of the invention. Preferably the polynucleotides are isolated free of other cellular components such as membrane components, proteins, and lipids. They can be made by a cell and isolated, or synthesized using an amplification technique, such as PCR, or by using an automatic synthesizer. Methods for purifying and isolating polynucleotides are routine and are known in the art.

Isolated polynucleotides also include oligonucleotide probes, which comprise at least one of the sequences shown in SEQ ID NOS:1-1563. An oligonucleotide probe is preferably at least 10, 11, 12, 13, 14, 15, 20, 30, 40, or 50 or more nucleotides in length. If desired, a single oligonucleotide probe can comprise 2, 3, 4, or 5 or more of the sequences shown in SEQ ID NOS:1-1563. The probes may or may not be labeled. They may be used, for example, as primers for amplification reactions , such as PCR, in Southern or Northern blots, or for *in situ* hybridization.

Oligonucleotide probes of the invention can be made by expressing cDNA molecules comprising one or more of the sequences shown in SEQ ID NOS:1-1563 in an expression vector in an appropriate host cell. Alternatively, oligonucleotide probes can be synthesized chemically, for example using an automated oligonucleotide synthesizer, as is known in the art.

Solid Supports Comprising Polynucleotides

Polynucleotides, particularly oligonucleotide probes, preferably are immobilized on a solid support. A solid support can be any surface to which a polynucleotide can be attached. Suitable solid supports include, but are not limited to, glass or plastic slides, tissue culture plates, microtiter wells, tubes, gene "chips,"or particles such as beads, including but not limited to latex, polystyrene, or glass beads. Any method known in the art can be used to attach a polynucleotide to a solid support, including use of covalent and non-covalent linkages, passive absorption, or pairs of binding moieties attached respectively to the polynucleotide and the solid support.

Polynucleotides are preferably present on an array so that multiple polynucleotides can be simultaneously tested for hybridization to polynucleotides present in a single biological sample. The polynucleotides can be spotted onto the array or synthesized *in situ* on the array. Such methods include older technologies, such as "dot blot" and "slot blot" hybridization (**53, 54**), as well as newer "microarray" technologies (**55-58**). A single array contains at least one polynucleotide, but can contain more than 100, 500, 1,000, 10,000, or 100,000 or more different probes in discrete locations.

Determining expression of a gene product

Each of the methods of the invention involves measuring expression of a gene product of at least one of the genes identified in Tables 3, 4, and 5 (SEQ ID NOS:1-1448). If desired, expression of gene products of at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 50, 75, 100, 125, 250, 500, 1,000, 1,250, or more genes can be determined.

Either protein or RNA products of the disclosed genes can be determined. Either qualitative or quantitative methods can be used. The presence of protein products of the disclosed genes can be determined, for example, using a variety of techniques known to the art, including immunochemical methods such as radioimmunoassay, Western blotting, and immunohistochemistry. Alternatively, protein synthesis can be determined *in vivo*, in a cell culture, or in an *in vitro* translation system by detecting incorporation of labeled amino acids into protein products.

RNA expression can be determined, for example, using at least 1, 2, 3, 4, 5, 10, 15,

20, 25, 30, 50, 75, 100, 125, 250, 500, 1,000, 5,000, 10,000, or 100,000 or more oligonucleotide probes, either in solution or immobilized on a solid support, as described above. Expression of the disclosed genes is preferably determined using an array of oligonucleotide probes immobilized on a solid support. *In situ* hybridization can also be used to detect RNA expression.

Identification of Cell Types

Cell-type specific genes are expressed at a level greater than 10 copies per cell in a particular cell type, such as epithelial cells of the colon, breast, lung, and kidney, keratinocytes, melanocytes, and cells from the prostate and brain, but are not expressed in cells of other tissues. Such cell-type specific genes represent "cell-type specific transcriptomes." The fraction of cell-type-specific transcripts ranges from 0.05% in normal prostate to 1.76% in normal colon epithelium. Approximately 50% of these transcripts tags match known genes or ESTs. The vast majority of these cell-type-specific genes have not been previously reported in the literature to be cell-type specific.

Cell type-specific genes are shown in Table 3. Genes which comprise the sequences shown in SEQ ID NOS:1-85 are uniquely expressed in colon epithelial cells. Genes which comprise the sequences shown in SEQ ID NOS:86-151 are uniquely expressed in brain cells. Genes which comprise the sequences shown in SEQ ID NOS:152-155 are uniquely expressed in keratinocytes. Genes which comprise the sequences shown in SEQ ID NOS:156-160 are uniquely expressed in breast epithelial cells. Genes which comprises the sequences shown in SEQ ID NOS:161-167 are uniquely expressed in lung epithelial cells. Genes which comprises the sequences shown in SEQ ID NOS:168-208 are uniquely expressed in melanocytes. Genes which comprise the sequences shown in SEQ ID NOS:209 and 210 are uniquely expressed in prostate cells. Genes which comprise the sequences shown in SEQ ID NOS:211-225 are uniquely expressed in kidney epithelial cells. Thus, determination of expression of at least one gene from each of these uniquely expressed groups, particularly those not previously known to be uniquely expressed, can be used to identify a test cell as an epithelial cell of the colon, breast, lung, and kidney, a keratinocyte, a melanocyte, or a cell from the prostate or brain.

Test cells can be obtained, for example, from biopsy or surgical samples, forensic samples, cell lines, or primary cell cultures. Test cells include normal as well as cancer cells, such as primary or metastatic cancer cells.

To identify a test cell as an epithelial cell of the colon, breast, lung, and kidney, a keratinocyte, a melanocyte, or a cell from the prostate or brain, expression of a gene product of at least one gene is determined, using methods such as those described above. If a test cell expresses a gene comprising a sequence shown in SEQ ID NOS:2, 5-18, and 20-85, the test cell is identified as a colon epithelial cell. If a test cell expresses a gene comprising a sequence shown in SEQ ID NOS:87-96, 98, 100-103, 105, 107-110, 112-129, and 131-151, the test cell is identified as a brain cell. If a test cell expresses a gene comprising a sequence shown in SEQ ID NOS:152-155, the test cell is identified as a keratinocyte. If a test cell expresses a gene comprising a sequence shown in SEQ ID NOS:156-160, the test cell is identified as a breast epithelial cell. If a test cell expresses a gene comprising a sequence shown in SEQ ID NOS:161-167, the test cell is identified as a lung epithelial cell. Expression of a gene comprising a sequence shown in SEQ ID NOS:168, 170, 172-177, 179-188, and 190-208 identifies the test cell as a melanocyte. Expression of a gene comprising a sequence shown in SEQ ID NOS:209 and 210 identifies the test cell as a prostate cell. Expression of a gene which comprises a sequence shown in SEQ ID NOS:211-225 identifies the test cell as a kidney epithelial cell.

Identifying a Test Cell as a Cancer Cell

A cancer-specific gene is expressed at a level of at least 3 copies per cancer cell, such as a colon cancer, breast cancer, brain cancer, lung cancer, or melanoma cell, at a level which is at least two-fold higher than expression of the same gene in a corresponding normal cell. Cancer-specific genes which comprise the sequences shown in SEQ ID NOS:226-265 (Table 5) represent a "cancer transcriptome." SEQ ID NOS:237, 239, 257, and 263 are sequences which are found in transcripts of novel cancer-specific genes of the invention. Oligonucleotide probes corresponding to cancer-specific genes can be used, for example, to detect and/or measure expression of cancer-specific genes for diagnostic purposes, to assess efficacy of various treatment regimens, and to screen for potential anti-

cancer drugs.

For example, determination of the expression level of any of these genes in a test cell relative to the expression level of the same gene in a normal cell (a cell which is known not to be a cancer cell) can be used to determine whether the test cell is a cancer cell or a non-cancer cell.

Test cells can be any human cell suspected of being a cancer cell, including but not limited to a colon epithelial cell, a breast epithelial cell, a lung epithelial cell, a kidney epithelial cell, a melanocyte, a prostate cell, and a brain cell. Test cells can be obtained, for example, from biopsy samples, surgically excised tissues, forensic samples, cell lines, or primary cell cultures. Comparison can be made to a non-cancer cell type, including to the corresponding non-cancer cell type, either at the time expression is measured in the test cell or by reference to a previously determined expression standard.

To identify a test cell as a cancer cell, expression of a gene product of at least one gene is determined, using methods such as those described above. The at least one gene comprises a sequence selected from the group consisting of SEQ ID NOS:226-265, particularly from the group consisting of SEQ ID NOS:228, 230-236, 238, 240-256, 258-260, and 262-265. An increase in expression of the at least one gene in the test cell which is at least two-fold more than the expression of the at least one gene in a cell which is not cancerous identifies the test cell as a cancer cell.

Reducing Cancer-Specific Gene Expression

Cancer-specific genes provide potential therapeutic targets for treating cancer or for use in model systems, for example, to screen for agents which will enhance the effect of a particular compound on a potential therapeutic target. Thus, a reagent can be administered to a human cell, either *in vitro* or *in vivo*, to reduce expression of a cancer-specific gene. The reagent specifically binds to an expression product of a gene comprising a sequence selected from the group consisting of SEQ ID NOS:226-265, particularly from the group consisting of SEQ ID NOS:228, 230-236, 238, 240-256, 258-260, and 262-265.

If the expression product is a protein, the reagent is preferably an antibody. Protein products of cancer-specific genes can be used as immunogens to generate antibodies, such

as a polyclonal, monoclonal, or single-chain antibodies, as is known in the art. Protein products of cancer-specific genes can be isolated from primary or metastatic tumors, such as primary colon adenocarcinomas, lung cancers, astrocytomas, glioblastomas, breast cancers, and melanomas. Alternatively, protein products can be prepared from cancer cell lines such as SW480, HCT116, DLD1, HT29, RKO, 21-PT, MDA-468, A549, and the like. If desired, cancer-specific gene coding sequences can be expressed in a host cell or in an *in vitro* translation system. An antibody which specifically binds to a protein product of a cancer-specific gene provides a detection signal at least 5-, 10-, or 2-fold higher than a detection signal provided with other proteins when used in an immunochemical assay. Preferably, the antibody does not detect other proteins in immunochemical assays and can immunoprecipitate the cancer-specific protein product from solution.

For administration *in vitro*, an antibody can be added to a tissue culture preparation, either as a component of the medium or in addition to the medium. In another embodiment, antibodies are delivered to specific tissues *in vivo* using receptor-mediated targeted delivery. Receptor-mediated DNA delivery techniques are taught in, for example, Findeis *et al. Trends in Biotechnol. 11*, 202-05, (1993); Chiou *et al.*, GENE THERAPEUTICS: METHODS AND APPLICATIONS OF DIRECT GENE TRANSFER (J.A. Wolff, ed.) (1994); Wu & Wu, *J. Biol. Chem. 263*, 621-24, 1988; Wu *et al.*, *J. Biol. Chem. 269*, 542-46, 1994; Zenke *et al.*, *Proc. Natl. Acad. Sci. U.S.A. 87*, 3655-59, 1990; Wu *et al.*, *J. Biol. Chem. 266*, 338-42, 1991.

If single-chain antibodies are used, polynucleotides encoding the antibodies can be constructed and introduced into cells using well-established techniques including, but not limited to, transferrin-polycation-mediated DNA transfer, transfection with naked or encapsulated nucleic acids, liposome-mediated cellular fusion, intracellular transportation of DNA-coated latex beads, protoplast fusion, viral infection, electroporation, "gene gun," and DEAE- or calcium phosphate-mediated transfection.

Effective *in vivo* dosages of an antibody are in the range of about 5 μg to about 50 μg/kg of patient body weight, about 50 μg to about 5 mg/kg, about 100 μg to about 500 μg/kg of patient body weight, and about 200 to about 250 μg/kg. For administration of polynucleotides encoding single-chain antibodies, effective *in vivo* dosages are in the range

of about 100 ng to about 200 ng, 500 ng to about 50 mg, about 1 μg to about 2 mg, about 5 μg to about 500 μg, and about 20 μg to about 100 μg of DNA.

If the expression product is mRNA, the reagent is preferably an antisense oligonucleotide. The nucleotide sequence of an antisense oligonucleotide is complementary to at least a portion of the sequence of the cancer-specific gene. Preferably, the antisense oligonucleotide sequence is at least 10 nucleotides in length, but can be at least 11, 12, 15, 20, 25, 30, 35, 40, 45, or 50 or more nucleotides long. Longer sequences also can be used. An antisense oligonucleotide which specifically binds to an mRNA product of a cancer-specific gene preferably hybridizes with no more than 3 or 2 mismatches, preferably with no more than 1 mismatch, even more preferably with no mismatches.

Antisense oligonucleotides can be deoxyribonucleotides, ribonucleotides, or a combination of both. Oligonucleotides, including modified oligonucleotides, can be prepared by methods well known in the art (47-52) and introduced into human cells using techniques such as those described above. The cells can be in a primary culture of human tumor cells, in a human tumor cell line, or can be primary or metastatic tumor cells present in a human body.

Preferably, a reagent reduces expression of a cancer-specific gene by at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, or 80% relative to expression of the gene in the absence of the reagent. Most preferably, the level of gene expression is decreased by at least 90%, 95%, 99%, or 100%. The effectiveness of the mechanism chosen to decrease the level of expression of a cancer-specific gene can be assessed using methods well known in the art, such as hybridization of nucleotide probes to cancer-specific gene mRNA, quantitative RT-PCR, or immunologic detection of a protein product of the cancer-specific gene.

Screening for Anti-Cancer Drugs

According to the invention, test compounds can be screened for potential use as anti-cancer drugs by assessing their ability to suppress or decrease the expression of at least one cancer-specific gene. The cancer-specific gene comprises a sequence selected from the group consisting of SEQ ID NOS:226-265, particularly from the group consisting of SEQ

ID NOS:228, 230-236, 238, 240-256, 258-260, and 262-265. Test compounds can be pharmacologic agents already known in the art or can be compounds previously unknown to have any pharmacological activity, including small molecules from compound libraries. Test substances can be naturally occurring or designed in the laboratory. They can be isolated from microorganisms, animals, or plants, or can be produced recombinantly or synthesized by chemical methods known in the art.

To screen a test compound for use as a possible anti-cancer drug, a cancer cell is contacted with the test compound. The cancer cell can be a cell of a primary or metastatic tumor, such as a tumor of the colon, breast, lung, prostate, brain, or kidney, or a melanoma, which is isolated from a patient. Alternatively, a cancer cell line, such as colon cancer cell lines HCT116, DLD1, HT29, Caco2, SW837, SW480, and RKO, breast cancer cell lines 21-PT, 21-MT, MDA-468, SK-BR3, and BT-474, the A549 lung cancer cell line, and the H392 glioblastoma cell line, can be used.

Expression of a gene product of at least one gene is determined using methods such as those described above. The gene comprises a sequence selected from the group consisting of SEQ ID NOS:226-265, preferably from the group consisting of SEQ ID NOS:228, 230-236, 238, 240-256, 258-260, and 262-265, even more preferably from the group consisting of SEQ ID NOS:237, 239, 257, and 263. A decrease in expression of the gene in the cancer cell identifies the test compound as a potential anti-cancer drug.

Standardizing Expression of a Test Gene

Genes which comprise the sequences shown in SEQ ID NOS:266-1448 (Table 4) are expressed at a level of at least five transcript copies per cell in every cell type analyzed, including epithelia of the colon, breast, lung, and kidney, melanocytes, chondrocytes, monocytes, cardiomyocytes, keratinocytes, prostate cells, and astrocytes, oligodendrocytes, and other cells present in the white matter of brain. These genes thus represent members of the "minimal transcriptome," the set of genes expressed in all human cells. The minimal transcriptome includes well known genes which are often used as experimental controls to normalize gene expression, such as glyceraldehyde 3-phosphate dehydrogenase, elongation factor 1 alpha, and gamma actin.

Ubiquitously expressed genes can be used to compare expression of a test gene in a test sample to expression of a gene in a standard sample. A ubiquitously expressed gene preferably comprises a sequence shown in SEQ ID NOS:266-375, 377-652, 654-796, and 798-1448, and more preferably comprises a sequence shown in SEQ ID NOS:282, 288, 300, 302, 308, 320, 323, 363, 368, 379, 381, 444, 453, 518, 531, 535, 538, 542, 579, 580, 594, 600, 604, 617, 626, 641, 650, 717, 728, 776, 777, 794, 818, 822, 842, 885, 887, 899, 900, 902, 904, 914, 930, 960, 964, 1001, 1015, 1020, 1027, 1035, 1090, 1113, 1119, 1146, 1151, 1163, 1233, 1235, 1252, 1255, 1270, 1340, 1345, 1356, 1359, 1360, 1362, 1385, 1415, and 1441.

Two ratios are determined using gene expression assays such as those described above. The first ratio is an amount of an expression product of a test gene in a test sample to an amount of an expression product of at least one ubiquitously expressed gene comprising a sequence selected from the group consisting of SEQ ID NOS:266-375, 377-652, 798-1447, and 1448 in the test sample. The second ratio is an amount of an expression product of the test gene in a standard sample to an amount of an expression product of the ubiquitously expressed gene in the standard sample. Expression of either the test gene or the ubiquitously expressed gene can be used as the denominator. If desired, multiple ratios can be determined, such as (a) an amount of an expression product of more than one test gene to that of a single ubiquitously expressed gene, (b) an amount of an expression product of a single test gene to that of more than one ubiquitously expressed genes, or (c) an amount of an expression product of more than one test gene to that of more than one ubiquitously expressed gene. Optionally, the ratio in the standard sample can be pre-determined.

The ratios determined in the test and standard samples are compared. A different between the ratios indicates a difference in the amount of the expression product of the test gene in the test sample.

The standard and test samples can be matched samples, such as whole cell cultures or homogenates of cells (such as a biopsy sample) and differ only in that the test biological sample has been subjected to a different environmental condition, such as a test compound, a drug whose effect is known or unknown, or altered temperature or other environmental

condition. Alternatively, the test and standard samples can be corresponding cell types which differ according to developmental age. In one embodiment, the test sample is a cancer cell, such as a colon cancer, breast cancer, lung cancer, melanoma, or brain cancer cell, and the standard sample is a normal cell.

The test gene can be a gene which encodes a protein whose biological function is known or unknown. Preferably the ratio of expression between the test gene and expression of the ubiquitously expressed gene is consistent in the standard sample. Even more preferably, expression of the ubiquitously expressed gene is not altered in the test sample. A difference between the first ratio of expression in the test sample and a second ratio of expression in the standard sample can therefore be used to indicate a difference in expression of the test gene in the test sample.

Screening for Compounds for Increasing an Organ or Cell Function

Test compounds can be screened for the ability to increase an organ or cell function by assessing their ability to increase expression of at least one tissue-specific gene. The tissue-specific gene comprises a sequence selected from at least one of the following groups:

(a) the sequences shown in SEQ ID NOS:2, 5-18, 20-84, and 85;

(b) the sequences shown in SEQ ID NOS:87-96, 98, 100-103, 105, 107-110, 112-129, 131-150, and 151;

(c) the sequences shown in SEQ ID NOS:152-154, and 155;

(d) the sequences shown in SEQ ID NOS:156-159 and 160;

(e) the sequences shown in SEQ ID NOS:161-166 and 167;

(f) the sequences shown in SEQ ID NOS:168, 170, 172-177, 179-188, 190-207, and 208;

(g) the sequences shown in SEQ ID NOS:209 and 210; and

(h) the sequences shown in SEQ ID NOS:211-224 and 225.

As with the anti-cancer drug screening method described above, test compounds can be pharmacologic agents already known in the art or can be compounds previously unknown to have any pharmacological activity, including small molecules from compound libraries.

Test substances can be naturally occurring or designed in the laboratory. They can be isolated from microorganisms, animals, or plants, or can be produced recombinantly or synthesized by chemical methods known in the art.

To screen a test compound for the ability to increase an organ or cell function, a cell, such as a colon epithelial cell, a brain cell, a keratinocyte, a breast epithelial cell, a lung epithelial cell, a melanocyte, a prostate cell, or a kidney cell, is contacted with the test compound. The cell can be a primary culture, such as an explant culture, of tissue obtained from a human, or can originate from an established cell line.

Expression of a gene product of at least one gene is determined using methods such as those described above. An increase in expression of a gene product of at least one gene comprising a sequence selected from (a) identifies the test compound as a potential drug for increasing a function of a colon cell. An increase in expression of a gene product of at least one gene comprising a sequence selected from (b) identifies the test compound as a potential drug for increasing a function of a brain cell. An increase in expression of a gene product of at least one gene comprising a sequence selected from (c) identifies the test compound as a potential drug for increasing a function of a skin cell. An increase in expression of a gene product of at least one gene comprising a sequence selected from (d) identifies the test compound as a potential drug for increasing a function of a breast cell. An increase in expression of a gene product of at least one gene comprising a sequence selected from (e) identifies the test compound as a potential drug for increasing a function of a lung cell. An increase in expression of a gene product of at least one gene comprising a sequence selected from (f) identifies the test compound as a potential drug for increasing a function of a melanocyte. An increase in expression of a gene product of at least one gene comprising a sequence selected from (g) identifies the test compound as a potential drug for increasing a function of a prostate cell. An increase in expression of a gene product of at least one gene comprising a sequence selected from (h) identifies the test compound as a potential drug for increasing a function of a kidney cell.

Restoring Function to a Diseased Tissue or Cell

Function can be restored to a diseased tissue or cell, such as a melanocyte or a colon,

brain, keratinocyte, breast, lung, prostate, or kidney cell, by delivering an appropriate tissue-specific gene to cells of that tissue. The tissue specific gene comprises a nucleotide sequence selected from at least one of the following groups:

(a) the sequences shown in SEQ ID NOS:2, 5-18, 20-84, and 85 (colon-specific);

(b) the sequences shown in SEQ ID NOS:87-96, 98, 100-103, 105, 107-110, 112-129, 131-150, and 151 (brain-specific);

(c) the sequences shown in SEQ ID NOS:152-154, and 155 (keratinocyte-specific);

(d) the sequences shown in SEQ ID NOS:156-159 and 160 (breast-specific);

(e) the sequences shown in SEQ ID NOS:161-166 and 167 (lung-specific);

(f) the sequences shown in SEQ ID NOS:168, 170, 172-177, 179-188, 190-207, and 208 (melanocyte-specific);

(g) the sequences shown in SEQ ID NOS:209 and 210 (prostate-specific); and

(h) the sequences shown in SEQ ID NOS:211-224 and 225 (kidney-specific).

Expression of the gene in a cell of the diseased tissue preferably is 10, 20, 30, 40, 50, 60, 70, 80, or 90% less than expression of the gene in a cell of the corresponding tissue which is normal. In some cases, the diseased cell fails to express the gene. A tissue-specific gene which is administered to cells for this purpose includes a polynucleotide comprising a coding sequence which is intron-free, such as a cDNA, as well as a polynucleotide which comprises elements in addition to the coding sequence, such as regulatory elements.

Coding sequences of many of the tissue-specific genes disclosed herein are publicly available. For the novel tissue-specific genes identified here, coding sequences can be obtained using a variety of methods, such as restriction-site PCR (Sarkar, *PCR Methods Applic. 2*:318-322, 1993), inverse PCR (Triglia *et al.*, *Nucleic Acids Res. 16*:8186, 1988), capture PCR (Lagerstrom, *et al.*, *PCR Methods Applic. 1*:111-119, 1991). Alternatively, the partial sequences disclosed herein can be nick-translated or end-labeled with ^{32}P using polynucleotide kinase using labeling methods known to those with skill in the art (BASIC METHODS IN MOLECULAR BIOLOGY, Davis *et al.*, eds., Elsevier Press, N.Y., 1986). A lambda library prepared from the appropriate human tissue can then be directly screened with the labelled sequences of interest.

Many methods for introducing polynucleotides into cells or tissues are available and

can be used to deliver a tissue-specific gene to a cell *in vitro* or *in vivo*. Introduction of the tissue-specific gene into a cell can be accomplished by any method by which a nucleic acid molecule can be inserted into a cell, such as transfection, electroporation, microinjection, lipofection, adsorption, and protoplast fusion. For *in vitro* administration, a tissue-specific gene can be added to a tissue culture preparation, either as a component of the medium or in addition to the medium. *In vivo* administration can be by means of direct injection of a vector comprising a tissue-specific gene to the particular tissue or cells to which the tissue-specific gene is to be delivered. Alternatively, the tissue-specific gene can be included in a vector which is capable of targeting a particular tissue and administered systemically (59-61).

For *in vitro* administration, suitable concentrations of a tissue-specific gene in the culture medium range from at least about 10 pg to 100 pg/ml, about 100 pg to about 500 pg/ml, about 500 pg to about 1 ng/ml, about 1 ng to about 10 ng/ml, about 10 ng to about 100 ng/ml, or about 100 ng/ml to about 500 ng/ml. For local administration, effective dosages of a tissue-specific gene range from at least about 10 ng to about 100 ng, about 50 ng to 150 ng, about 100 ng to about 250 ng, about 1 μg to about 10 μg, about 5 μg to about 50 μg, about 25 μg to about 100 μg, about 75 μg to about 250 μg, about 100 μg to about 250 μg, about 200 μg to about 500 μg, about 500 μg to about 1 mg, about 1 mg to about 10 mg, about 5 mg to about 50 mg, about 25 mg to about 100 mg, or about 50 mg to about 200 mg of DNA per injection. Suitable concentrations for systemic administration range from at least about 500 ng to about 50 mg, about 1 μg to about 2 mg, about 5 μg to about 500 μg, and about 20 μg to about 100 μg of DNA per kg of body weight.

Recombinant DNA technologies can be used to improve expression of the tissue-specific gene by manipulating, for example, the number of copies of the gene in the cell, the efficiency with which the gene is transcribed, the efficiency with which the resultant transcripts are translated, and the efficiency of post-translational modifications. Recombinant techniques useful for increasing the expression of a tissue-specific gene in a cell include, but are not limited to, providing the tissue-specific gene in a high-copy number plasmid, integrating the tissue-specific gene into one or more host cell chromosomes, adding vector stability sequences to plasmids, substituting or modifying

transcription control signals (*e.g.*, promoters, operators, enhancers), substituting or modulating translational control signals (*e.g.*, ribosome binding sites, Shine-Dalgarno sequences), and deleting sequences that destabilize transcripts. (*See* Dow *et al.*, U.S. Patent 5,935,568).

Preferably, delivery of the tissue-specific gene increases expression of a gene product of the tissue-specific gene in the cell or tissue by at least 10, 20, 30, 40, 50, 60 70, 80, 90, 95, 98, 99, or 100% relative to expression of the tissue-specific gene in a diseased cell or tissue to which the gene has not been delivered. Expression of a protein product of the tissue-specific gene can be determined immunologically, using methods such as radioimmunoassay, Western blotting, and immunohistochemistry. Alternatively, incorporation of labeled amino acids into a protein product can be determined. RNA expression is preferably determined using one or more oligonucleotide probes, either in solution or immobilized on a solid support, as described above.

All documents cited in this disclosure are expressly incorporated herein. The above disclosure generally describes the present invention, and all references cited in this disclosure are incorporated by reference herein. A more complete understanding can be obtained by reference to the following specific examples which are provided for purposes of illustration only and are not intended to limit the scope of the invention.

EXAMPLE 1

Tissue samples and the SAGE method

RNA for normal tissues was obtained from the following sources: colon epithelial cells isolated from sections of normal colon mucosa from two patients (**41**); HaCaT keratinocyte cells (**42**), normal mammary epithelial cells from two individuals (Clonetics); normal bronchial epithelial cell from two individuals (**43**); normal melanocytes from two individuals (Cascade Biologics); normal cultured monocytes, dendritic cells and TNF activated dendritic cells; two normal kidney epithelial cell lines; cultured chondrocyte cells from two normal individuals and one patient with osteoarthritic disease; normal fetal cardiomyocytes in normoxic and hypoxic conditions; and normal brain white matter from

two patients and normal cultured astrocyte cells.

RNA for diseased tissues was obtained from the following sources: primary colon adenocarcinomas from two patients, HCT116, DLD1, HT29, Caco2, SW837, SW480, and RKO colon cancer cell lines cultured *in vitro* in a variety of different cellular conditions including log phase growth, G1/G2 phase growth arrest, and apoptosis (**40, 41, 44, 45**); primary pancreatic adenocarcinomas from two patients and ASPC-1 and PL-45 pancreatic cancer cell lines (**41**); breast cancer cell lines 21-PT, 21-MT, MDA-468, SK-BR3, and BT-474; primary lung squamous cell cancers from two patients (**43**), primary lung adenocarcinoma from one patient, and the A549 lung cancer cell line (**43**); primary melanomas from 3 patients; kidney epithelial cells lines from two patients with polycystic kidney disease; hemangiopericytomas from 5 patients; primary glioblastoma tumors from two patients; and the H392 glioblastoma cell line.

Isolation of polyadenylate RNA and the SAGE method for all tissues was performed as previously described (**1, 12**; *see also* U.S. Patents 5,866,330 and 5,695,937).

EXAMPLE 2

Data analysis

The SAGE software (**12**) was used to analyze raw sequence data and to identify a total of 3,668,175 SAGE tags. Of these, 171,346 tags (4.7%) corresponded to linker sequences and were removed from further analysis. The remaining 3,496,829 tags were derived from transcript sequences, but a small fraction of these contained sequencing errors. SAGE analysis of yeast (**1**), for which the entire genome sequence is known, demonstrated a sequencing error rate of ~0.7% per bp, translating to a tag error rate of 6.8% (1-0.993; **10**), in accord with sequence errors measured in the current data set.

To provide as accurate an estimate of unique genes as possible, we accounted for sequencing errors in two ways. First, we only considered tags that occurred twice in the data set. Although this requirement might have removed legitimate transcript tags expressed at very low levels (less than approximately 0.2 copies per cell, or 2 copies in 3,496,829 transcript tags), it eliminated the majority of sequencing errors (172,276 tags).

Second, because of the size of the data set utilized, it was possible that the same

sequencing error in a given tag may be observed multiple times. To account for these, tags with expression levels high enough to give multiple redundant errors were analyzed for single base substitutions, insertions, and deletions. If the observed expression level of a tag did not exceed its expected incidence due to redundant errors by a factor of five, it was assumed to be the result of a repeated sequencing error. This identified and removed an additional 27,051 unique tags (156,174 total tags), a number very similar to estimates of multiple sequencing errors obtained by Monte Carlo simulations.

In total, these corrections amount to a sequencing error rate of approximately 9.4%, suggesting that our analyses more than fully accounted for sequencing errors and that the remaining 134,135 unique transcript tags represented a conservative accounting of legitimate transcripts.

Transcript tags were matched to known genes and ESTs by use of tables containing matching 10 bp transcript sequences, UniGene clusters, GenBank accession numbers, and functional descriptions downloaded from the SAGEmap web site (http://www.ncbi.nlm.nih.gov/SAGE) (Lal *et al.*, in press) on Feb 23, 1999 (UniGene build 70, http://www.ncbi.nlm.nih.gov/UniGene), and the Microsoft Access software. As UniGene clusters numbers may change over time, the most recent tag to cluster mapping can be obtained for each transcript tag individually at http://www.ncbi.nlm.nih.gov/SAGE/SAGEtag.cgi, or for the entire data set at http://www.sagenet.org./transcriptome. A total of 37,534 distinct transcripts from the UniGene database contained polyadenylation signals or polyadenylated tails and matched the collection of SAGE transcript tags; these corresponded to 23,534 unique UniGene clusters.

Transcript abundance per cell was determined simply by dividing the observed number of tags for a given transcript by the total number of transcripts obtained. An estimate of about 300,000 transcripts per cell was used to convert the abundances to copies per cell (**46**). For tissue specific transcripts, only transcript tags expressed at nominally ≥ 10 transcript copies per cell were considered in order to normalize for tissues with fewer total tags analyzed.

The following transcript data from this analysis are available electronically at the

SAGEnet web site (http://www.sagenet.org/transcriptome) with the corresponding expression levels and UniGene descriptions: 134,135 unique transcript tags identified from 3.5 million total transcripts tags; 69,381 transcript tags identified from colon cancer cells; 217 transcripts that are exclusively expressed in colon epithelium, keratinocytes, breast epithelium, lung epithelium, melanocytes, kidney epithelium and cells from prostate and brain; 987 transcripts that were expressed in all tissues. Individual transcript libraries from a total of ~800,000 transcript tags from colon epithelium, normal brain, colon cancer, and brain cancer are available at the SAGEmap web site (http://www.ncbi.nlm.nih.gov/SAGE) (Lal *et al.*, in press).

EXAMPLE 3

Estimation of the number of genes present in the human genome

The transcripts detected by SAGE provides an estimate of the number of genes present in the human genome. Historically, estimates of the number of unique genes in the genome have ranged from 60,000 to over 100,000 genes using analyses of EST clustering (**15**), frequency of genes in characterized genomic regions, frequency of CpG islands (**16**), and RNA-cDNA reassociation kinetics (**4**). If one were to assume that each unique transcript tag observed by SAGE corresponded to a unique gene, our data would indicate that there are approximately 134,000 genes in the human genome.

However, such an approach is likely to overestimate the number of unique genes in the genome, as distinct transcripts can be derived from a single gene. Multiple sites for polyadenylation (**17**), alternative splicing, premature transcriptional termination (**18**), as well as polymorphisms in the SAGE tag or nearby restriction endonuclease site could lead to multiple transcript tags for any one gene. An analysis of all publicly available 3' end-derived ESTs revealed that this was the case for many transcripts, and provided an estimate of the multiplicity of transcripts expected for individual genes. 37,534 distinct 3' transcripts containing polyadenylation signals or polyadenylated tails were observed to correspond to 23,534 unique UniGene clusters, an average 1.6 different transcripts per gene. Applying a similar calculation to our SAGE data would suggest that the 134,135 transcripts observed corresponded to 84,103 unique genes. As our SAGE data is by no

means a complete analysis of transcripts from all possible tissues, this estimate would provide a lower boundary for the number of unique genes in the genome. This figure is significantly higher than the 65,538 genes estimated from a clustering of 982,808 ESTs (UniGene Build 70) (15), and suggests that a substantial number of genes expressed at low levels may not be present in current EST databases.

EXAMPLE 4

Assessment of transcriptome complexity

Assessment of transcriptome complexity requires a relatively complete sampling of a transcriptome for the cell type under analysis. Human cells are thought to contain close to 300,000 mRNA molecules, and therefore an analysis of at least several hundred thousand transcripts would be needed. Approximately 350,000 and 300,000 transcripts were analyzed from DLD1 and HCT116 colorectal cancer cells, respectively. As these cancer cells are diploid, have similar genetic and phenotypic properties, and have very similar gene expression patterns (see below), transcript tags obtained from these cells were analyzed in combination as well as individually.

Analysis of either cell line afforded approximately a one fold coverage of the 300,000 mRNA molecules in a cell, while the combined set represented a two fold coverage even for mRNA molecules present at a single copy per cell. Measurement of ascertained new tags at increasing increments of tags indicated that the fraction of new transcripts from analysis of additional tags approached 0 at approximately 650,000 tags in the combined set (FIG. 1). This suggested that generation of further SAGE tags would yield few additional genes, and Monte Carlo simulations indicated that analysis of 643,283 tags would identify at least one tag for a given transcript 96% of the time if its expression level was at least two transcript copies per cell, and 83% of the time if its expression level was at least one transcript copy per cell.

The combined 643,283 transcript tags represented 69,381 unique transcripts, of which 44,174 corresponded to known genes or ESTs in the GenBank or UniGene databases while 25,207 represented previously undescribed transcripts (Table 2). Even when accounting for multiple unique transcripts per gene, these transcripts would represent at least 43,502

unique genes. This is substantially higher than the previous estimate of 15,000-25,000 expressed genes obtained by RNA-DNA reassociation kinetics in a variety of human cell types (4), and suggests that a significant fraction of the genome may be expressed in individual cell types. As the kinetics of reassociation of a particular class of RNA and cDNA may be affected by a number of experimental variables and may underestimate transcripts of low abundance (4), it is not surprising that our studies have detected a higher number of expressed genes than estimated by hybridization analysis in both human cells (Table 2) and yeast.

EXAMPLE 5

Expression levels of transcripts in colon cancer cells

Expression levels of transcripts in the colon cancer cell ranged from 0.5 to 2341 copies per cell. The 61 transcripts expressed at over 500 transcript copies per cell made up nearly 1/4 of the mRNA mass of the cell and the most highly expressed 623 genes accounted for ½ of the mRNA content. In contrast, the vast majority of unique transcripts were expressed at low levels, with just under 23% of the mRNA mass of the cell comprising 90% of the unique transcripts expressed (Table 2). A "virtual rot" analysis of the expressed transcripts identified a relatively continuous distribution of gene expression without markedly discrete abundance classes, similar to those observed in previous rot studies of human cancer cells (20) (FIG. 2).

The identities of the expressed genes reveal the diversity of expression of a human transcriptome (data available at http://www.sagenet.org./transcriptome). For example, highly expressed genes often encoded proteins important in protein synthesis, energy metabolism, cellular structure and certain tissue specific functions. Moderate and low abundance genes accounted for a multitude of cellular processes including protein modification enzymes, DNA replication machinery, cell surface receptors, components of signal transduction pathways and transcription factors as well as many other transcripts with currently unknown functions.

EXAMPLE 6

Differences in gene expression between different tissues

Differences in gene expression between different tissues may provide insights into the specialized processes underlying human physiology in normal and diseased states. In line with previous observations, overall gene expression patterns among the 19 different tissues analyzed were similar (examples in FIGS. 3A-3C). Changes in gene expression between physiologic states of a particular cell type or between patient samples of the same tissue were less than changes between cell types of different origins (FIGS. 3A-3C). Likewise, only a small fraction of transcripts was exclusively expressed in a particular normal or disease tissue. Detailed analyses of transcripts from epithelia of colon, breast, lung, and kidney, melanocytes, and cells from prostate and brain, identified transcripts that were nominally expressed at greater than 10 copies per cell in one tissue but not in any other tissue studied. The fraction of these tissue-specific transcripts ranged from 0.05% in normal prostate to 1.76% in normal colon epithelium (Table 3). Approximately 50% of these transcript tags matched known genes or ESTs (examples in Table 3 and data available at http://www.sagenet.org/transcriptome). Some of these transcripts identified genes already reported to be important for tissue specific processes. For example, brain specific transcripts such as GABA receptor, myelin basic protein, and synaptopodin are known to be important for synaptic transmission (21) formation and maintenance of the myelin sheath (22) and dendrite shape and motility (23), respectively. Likewise, guanylin/uroguanylin (24), carbonic anhydrase 1 (25), and CDX2 (26) are known to be expressed in colonic epithelium. 5,6-dihydroxyindole-2-carboxylic acid oxidase has been shown to have an important role for normal melanocyte pigment synthesis (27), while expression of MART-1 and melastatin may have clinical implications for melanoma patients (28, 29). However, the vast majority of the tissue specific transcripts observed have not been previously reported in the literature and their roles in the tissues examined remain to be elucidated.

EXAMPLE 7

Minimal transcriptome

Nearly 1000 transcripts were detected that were expressed at 5 transcript copies per cell in every cell type analyzed. These expressed genes represent a view into the "minimal transcriptome," the set of genes expressed in all human cells. Such genes, listed in order of their uniformity of expression in Table 4 (and available at http://www.sagenet.org./transcriptome), largely represent well known constitutive or housekeeping genes thought to provide the molecular machinery necessary for basic functions of cellular life (4). Genes involved in DNA, RNA, protein, lipid and oligosaccharide biosynthesis as well as in energy metabolism were among those observed. Additionally, genes from other functional classes including structural proteins (*e.g.* dystroglycan and myosin light chain), signaling molecules (*e.g.* 14-3-3 proteins and MAPKK2), proteins with compartmentalized functions (*e.g.* lysosome-associated membrane glycoprotein and ER lumen retaining protein receptor 1), cell surface receptors (*e.g.* FGF receptor and STRL22 G protein coupled receptor), proteins involved in intracellular transport (*e.g.* syntaxin and alpha SNAP), membrane transporters (*e.g.* Na^+/K^+ ATPase and mitochondrial F1/F0 ATPase), and enzymes involved in post-translational modification and protein degradation (*e.g.* kinases, phosphatases and proteasome components) were observed and were not previously known to be ubiquitously expressed. Well known genes often used as experimental controls such glyceraldehyde 3-phosphate dehydrogenase, elongation factor 1 alpha, and gamma actin were observed but varied in expression as much as 6 fold among different cell types.

EXAMPLE 8

Genes involved in tumorigenesis

Genes that are uniformly expressed in cancers but expressed at lower levels in normal tissues may turn out to be important for tumorigenesis, and demonstrate how gene expression patterns might be useful in the analysis of disease states. We detected 40 genes that were expressed in all cancer tissues examined at levels 3 transcript copies per cell and whose expression was at least 2-fold higher in each cancer compared to its corresponding

normal tissue (Table 5). Four of these transcripts had no matches to known genes and 15 matched ESTs with no known function. Several of the highly induced transcripts provided tantalizing clues about their roles in tumorigenesis. For example, S100A4 has been thought to play a role in late stage tumorigenesis as it is overexpressed in colorectal adenocarcinomas but not adenomas (30), and its induction can promote (while its inhibition can prevent) metastasis in tumor models. Midkine, a heparin-binding growth factor has been reported to be overexpressed in certain cancers (34), to transform cells *in vitro* (35), and to promote tumor angiogenesis *in vivo*. Finally, overexpression of survivin, an IAP apoptosis inhibitor (37) has been recently shown to predict shorter survival rates in colorectal cancer patients and may carry out its antiapoptotic functions as a mitotic spindle checkpoint factor (39). The observed elevated expression of such genes in many tumor types indicates a potentially general role for these genes in tumorigenesis and suggests they may be useful as diagnostic markers or targets for therapeutic intervention.

EXAMPLE 9

Estimate of gene number

The 134,135 distinct transcripts identified in this study, corresponding to approximately 84,103 unique genes, provided an estimate of gene number substantially higher than the recent estimate (~ 65,000 genes) derived from extant EST clusters. What could account for the difference between these estimates, considering that both are derived from sequencing of transcripts from similar cell types? One explanation is that the clustering estimate is based on the number of observed EST clusters (62,236) divided by a measure of the completeness of the EST database. The latter value is calculated as the fraction of "characterized" genes in GenBank that already have EST matches (~95%). The characterized genes in GenBank have been assumed to be representative of the rest of the genes in the human genome, but our SAGE data indicated that their average expression was more than 10 fold higher than the mean levels of gene expression. Similarly, the number of ESTs that were present in clusters with characterized genes was approximately 12 fold higher than clusters composed entirely of ESTs. Such highly expressed genes would be more likely to be represented in transcript databases, thereby leading to an overestimation

of the completeness of the EST databases, and an underestimation of the number of unique genes. Indeed, the number of UniGene clusters continues to grow as a greater diversity of tissues is analyzed through the Cancer Genome Anatomy Project, and as of the date of submission of this manuscript already exceeds the recent EST derived estimate (71,849 gene clusters in Build 80 versus 65,538 predicted from Build 70).

Like other genome-wide analyses, studies of human transcriptomes using SAGE have several potential limitations. First, a small number of transcripts would be expected to lack the restriction enzyme site required to produce the 14 bp tags, and would therefore not be detected by our analyses (**12**). Second, our study was limited to the 19 tissues analyzed. Genes uniquely expressed in other tissues would not have been detected, and accordingly, genes observed to be tissue specific in our studies may turn out to be expressed in other normal or disease states. Finally, identification of genes corresponding to specific tags is mainly based on large but incomplete databases of ESTs and characterized genes. SAGE tags without matches to existing databases can directly be used to identify previously uncharacterized genes (**1, 12, 40**), but additional 3' EST data, as well as that of genomic regions would make gene identification more rapid.

REFERENCES

1. Velculescu *et al.*, *Cell* **88**, 243-251 (1997).
2. Pietu *et al.*, *Genome Res* **9** 195-209 (1999).
3. Wadman, *Nature* **398**, 177 (1999).
4. Lewin, *Gene Expression* **2**, 694-727 (1980).
5. Adams *et al.*, *Nature* **377**, 3 ff. (1995)
6. Okubo *et al.*, *DNA Res* **1**, 37-45 (1994).
7. Alwine *et al. Proc Natl Acad Sci USA* **74**, 5350-5354 (1977).
8. Zinn *et al. Cell* **34**, 865-879 (1983).
9. Veres *et al. Science* **237**, 415-417 (1987).
10. Hedrick *et al. Nature* **308**, 149-153 (1984).
11. Liang & Pardee, *Science* **257**, 967-971 (1992).
12. Velculescu *et al. Science* **270**, 484-487 (1995).
13. Kal *et al.*, Mol Biol Cell **10**, 1859-1872 (1999).
14. Basrai *et al.*, NORF5/HUG1 is a component of the MEC1 mediated checkpoint response to DNA damage and replication arrest in S. cerevisiae. *submitted.*
15. Fields *et al. Nat Genet* **7**, 345-346 (1994).
16. Antequera *et al. Proc Natl Acad Sci USA* **90** 11995-11999 (1993).
17. Gautheret *et al. Genome Res* **8**, 524-530 (1998).
18. Bouck *et al. Trends Genet* **15**, 159-62 (1999).
19. Bentley & Groudine, *Cell* **53**, 245-256 (1988).
20. Bishop *et al. Nature* **250**, 199-204 (1974).
21. Mody *et al. Trends Neurosci* **17**, 517-25 (1994).
22. Staugaitis *et al. Bioessays* **18**, 13-18 (1996).
23. Mundel *et al.*, *J Cell Biol* **139**, 193-204 (1997).
24. Wiegand *et al. FEBS Lett* **311**, 150-154 (1992).
25. Sowden *et al. Differentiation* **53**, 67-74 (1993).
26. Suh & Traber, *Mol Cell Biol* **16**, 619-625 (1996).
27. Blarzino *et al.*, *Free Radic Biol Med* **26**, 446-453 (1999).
28. Busam *et al. Adv Anat Pathol* **6**, 12-18 (1999).

29. Duncan *et al., Cancer Res* **58**, 1515-1520 (1998).
30. Takenage *et al., Clin Cancer Res* **3**, 2309-2316 (1997).
31. Lloyd *et al. Oncogene* **17**, 465-473 (1998).
32. Maelandsmo *et al., Cancer Res* **56**, 5490-5498 (1996).
33. Muramatsu & Muramatsu, *Biochem Biophy Res Commun* **177**, 652-658 (1991).
34. Tsutsui *et al., Cancer Res* **53**, 1281-1285 (1993).
35. Kadomatsu *et al., Br J Cancer* **75**, 354-359 (1997).
36. Choudhuri *et al. Cancer Res.* **57**, 1814-1819 (1997).
37. Ambrosini *et al. Nat Med* **3**, 917-921 (1997).
38. Kawasaki *et al., Cancer Res* **58**, 5071-5074 (1998).
39. Li *et al., Nature* **396**, 580-584 (1998).
40. Polyak *et al. Nature* **389**, 300-304 (1997).
41. Zhang *et al., Science* **276**, 1268-1272 (1997).
42. Boukam *et al., J Cell Biol* **106**, 761-771 (1988).
43. Hibi *et al., Cancer Res* **58**, 5690-5694 (1998).
44. Hermeking *et al., Molecular Cell* **1**, 3-11 (1997).
45. He *et al., Science* **281**, 1509-1512 (1998).
46. Hastie & Bishop, *Cell* **9**, 761-774 (1976).
47. Agrawal *et al., Trends Biotechnol.* **10**, 152-158 (1992)
48. Uhlmann *et al., Chem. Rev.* **90**, 543-584 (1990)
49. Uhlmann *et al., Tetrahedron. Lett.* **215**, 3539-3542 (1987)
50. Brown, *Meth. Mol. Biol.* **20**, 1-8 (1994)
51. Sonveaux, *Meth. Mol. Biol.* **26**, 1-72 (1994)
52. Uhlmann *et al., Chem. Rev.* **90**, 543-583 (1990)
53. White & Bancroft, *J. Biol. Chem.* **257**, 8569 (1982)
54. Sambrook *et al.*, MOLECULAR CLONING. A LABORATORY MANUAL, 2d ed., pages 7.53-7.57 (1989)
55. Chee *et al., Science* **274**, 610-14 (1996)
56. DeRisi *et al., Nat. Genet.* **14**, 457-60 (1996)
57. Schena, *Bioessays* **18**, 427-31 (1996)

58. Lockhart *et al.*, *Nature Biotechnology*, **14** (1996)

59. Romanczuk *et al.*, *Hum. Gene. Ther.* **10**, 2615-26

60. Lanzov, *Mol. Genet. Metab.* **68**, 276-82 (1999)

61. Lai & Lien, *Exp. Nephrol.* **7**, 11-14 (1999)

Table 1. Tissues and transcript tags analyzed

Normal tissues	Libraries	Total Transcripts	Unique Genes
Colon epithelium[1,2]	2	98,089	12,941
Keratinocytes[3]	2	83,835	12,598
Breast epithelium[3]	2	107,632	13,429
Lung epithelium[4]	2	111,848	11,636
Melanocytes[3]	2	110,631	14,824
Prostate[3]	2	98,010	9,786
Monocytes[3]	3	66,673	9,504
Kidney epithelium[3]	2	103,836	15,094
Chondrocytes[3]	4	88,875	11,628
Cardiomyocytes[3]	4	77,374	9,449
Brain[2]	3	202,448	23,580
Diseased Tissues			
Colon cancer[1,2,3]	22	1,004,509	56,153
Pancreatic cancer[1]	4	126,414	17,050
Breast cancer[3]	5	226,630	18,685
Lung cancer[4]	5	221,302	22,783
Melanoma[3]	10	269,332	25,600
Polycystic kidney disea	2	112,839	16,280
Hemangiopericytoma[3]	5	199,985	31,351
Brain cancer[2]	3	186,567	23,108
Total	84	3,496,829	84,103

1 Ref. 40, 41, 44, 45
2 Lal et al.
3 unpublished
4 Ref. 43

Table 2. Transcript abundance

Copies/Cell	Colon Cancer Cells	
	Unique transcripts	Mass fraction mRNA (%)
>500	61	20
Match GenBank (%)	61 (100)	
50 to 500	562	27
Match GenBank (%)	554 (99)	
5 to 50	6,358	30
Match GenBank (%)	6,023 (95)	
<=5	62,400	23
Match GenBank (%)	37,536 (60)	
Total	69,381	100
Match GenBank (%)	44,174 (64)	

Table 3. Tissue-specific genes

Tag sequence	SEQ ID NO:	Observed	Copies/cell	Unigene Description
Colon epithelium (1.76%)				
ATACTCCACT	1	141	431	Guanylate cyclase activator 2 (guanylin, intestinal, heat-stable)
TCAGCTGCAA	2	72	220	No match
GTCATCACCA	3	57	174	H.saplens mRNA for GCAP-II/uroguanylin precursor
CCTTCAAATC	4	46	141	Carbonic anhydrase I
ACACCCATCA	5	29	89	No match
CCAACACCAG	6	28	86	No match
AATAGTTTCC	7	23	70	Pregnancy-specific beta-1 glycoprotein 6
CCAGGCGTCA	8	18	55	No match
GAACAGCTCA	9	18	55	ESTs
TACTCGGCCA	10	15	46	No match
GGGGGAGAAG	11	12	37	ESTs
AGTGGGCTCA	12	11	34	No match
GAGCACCGTG	13	11	34	No match
GATCTATCCA	14	10	31	ESTs
GAACGCCAGA	15	9	28	No match
GCCCTCGGAG	16	9	28	ESTs
ACAAGCCTAG	17	9	28	No match
GTCACAGGAA	18	9	28	No match
GCCCTCGGAG	19	9	28	Human homeobox protein Cdx2 mRNA, complete cds
CTAGGATGAT	20	9	28	ESTs
CCAACTATCG	21	8	24	No match
CTGACGGGGA	22	8	24	ESTs
GAGGGTTTTA	23	8	24	Homo sapiens C19steroid specific UDP-glucuronosyltransferase mRNA, complete cds
GGGGTCCCAT	24	8	24	No match
GCCAGGTCAC	25	7	21	No match
AGAACACCAA	26	7	21	No match
AATCCCGCCC	27	7	21	Homo sapiens hAQP8 mRNA for aquaporin 8, complete cds
ACACTGCCTC	28	6	18	No match
AGAGTCCAGG	29	6	18	Homo sapiens carcinoembryonic antigen (CGM2) mRNA, complete cds
CCAGACGTAG	30	6	18	No match
GAGGCCCCCG	31	6	18	No match
CTGTGTGCCC	32	5	15	ESTs, Weakly similar to tryptase-III [H.sapiens]
GAGAGGATGG	33	5	15	ESTs
GGCTGAACCA	34	5	15	No match
CCAAATCATT	35	5	15	No match
ACGGCTGGGC	36	5	15	No match
ACCTTCATCT	37	5	15	EST
AGGGCTTGAG	38	5	15	No match
ACCTTCATCT	39	5	15	Human rearranged metabotropic glutamate receptor type II (GLUR2) mRNA, complete cds
TCAGGCCAGA	40	5	15	No match
CTGTGTGCCC	41	5	15	ESTs

Table 3, cont.

GGATGTCAAC	42	5	15	Human RecA-like protein (hREC2) mRNA, complete cds
ATCTGGAGCA	43	5	15	Alcohol dehydrogenase 1 (class I), alpha polypeptide
GAGAGGATGG	44	5	15	INTEGRAL MEMBRANE PROTEIN E16
ATCTGGAGCA	45	5	15	Alcohol dehydrogenase 3 (class I), gamma polypeptide
GGATGTCAAC	46	5	15	Polymeric immunoglobulin receptor
CACAGACACA	47	4	12	No match
TGCTCCTAAC	48	4	12	No match
TATACCCGGA	49	4	12	No match
TATCCTGATG	50	4	12	No match
GGCCCTCCCG	51	4	12	No match
GTAGCGATGG	52	4	12	Pim-1 oncogene
GCAGGTTGTG	53	4	12	No match
TGGGAACCGG	54	3	9	No match
ACACCTCTCT	55	3	9	No match
GGAAAACAGG	56	3	9	No match
CAGGCGGCAC	57	3	9	No match
CAGGTTGGTC	58	3	9	Homo sapiens hRVP1 mRNA for RVP1, complete cds
GGGATATAAA	59	3	9	No match
GTGGAAAATC	60	3	9	No match
GTGTGTGAAT	61	3	9	No match
ATGTGACACT	62	3	9	No match
ATGGTGTAAT	63	3	9	ESTs
TCACATTGAT	64	3	9	H.sapiens mRNA for LI-cadherin
TAACTAAACA	65	3	9	No match
TGCCCGGGTC	66	3	9	No match
TAGTCGGAAA	67	3	9	No match
GCTATACGGG	68	3	9	No match
TCACACCCCA	69	3	9	No match
CTGCCCGAAC	70	3	9	ESTs
AGTCACCTCT	71	3	9	No match
TCATTGGTTT	72	3	9	No match
TCCTCTCCTC	73	3	9	No match
CCTCTCGGCC	74	3	9	No match
CCACTGAAGT	75	3	9	No match
CTGGCTTGCT	76	3	9	No match
GAAAACAGAA	77	3	9	EST
AAAGCACGTC	78	3	9	No match
GAAAACAGAA	79	3	9	ESTs, Weakly similar to synapse-associated protein sap47-1 [D.melanogaster]
TTGATTCCAT	80	3	9	No match
AAACAGGCAC	81	3	9	No match
CTTACAGTCC	82	3	9	No match
GAATGGACTC	83	3	9	No match
GAACCCAAAC	84	3	9	No match
GAAAACAGAA	85	3	9	ESTs

Normal Brain (1.36)

Table 3, cont.

ACTTTGTCCC	86	160	237	Glial fibrillary acidic protein
GTGCGAATCC	87	79	117	ESTs
CAAAAAGTTA	88	36	53	ESTs
TTAACTTTAT	89	33	49	Homo sapiens neuroendocrine-specific protein A (NSP) mRNA, complete cds
CAGCCAAATG	90	29	43	ESTs
GCCTGTGGTG	91	28	41	Homo sapiens LY6H mRNA, complete cds
CTTAGGGACA	92	26	39	ESTs
TTGGAGGTGA	93	22	33	ESTs
ATTCCATTTC	94	20	30	ESTs
ATTCCATTTC	95	20	30	ESTs, Highly similar to RAS-RELATED PROTEIN RAB-10 [Canis familiaris]
AGAGAGCGGA	96	19	28	Human guanine nucleotide-binding regulatory protein (Go-alpha) gene
TTCTCAATAC	97	19	28	Homo sapiens mRNA for synaptopodin
CATCCTCCCA	98	19	28	No match
GTATCGATTT	99	16	24	Homo sapiens GABA-B receptor mRNA, complete cds
TTGTAAACAG	100	15	22	ESTs, Weakly similar to cyclin I [H.sapiens]
GCCCTGTATT	101	15	22	ESTs
CCACATTGCC	102	15	22	Homo sapiens chromosome 7q22 sequence
CAGGGCAACG	103	15	22	No match
AAAAGCAAAT	104	15	22	Human mRNA for MOBP (myelin-associated oligodendrocytic basic protein), complete cds, clone hOPRP1
ACCAATCCTA	105	14	21	Human guanine nucleotide-binding regulatory protein (Go-alpha) gene
CTGTGTGTCC	106	13	19	AXONIN-1 PRECURSOR
TCAGACAATA	107	12	18	ESTs
TGGTGAGATG	108	12	18	ESTs
ATTTTTTGTT	109	12	18	ESTs
ACATTGAGTC	110	12	18	Homo sapiens mRNA for MEGF4, partial cds
GTCAGTCTAC	111	11	16	Glutamate receptor, metabotropic 3
GTCCCACTTC	112	11	16	ESTs
GGGGCCCGAA	113	11	16	No match
TGACTCACCC	114	10	15	Homo sapiens calmodulin-stimulated phosphodiesterase PDE1B1 mRNA, complete cds
GACAGCGACA	115	10	15	No match
GGTGTACATA	116	10	15	ESTs
TAGCTATAAA	117	10	15	ESTs
GGTGTACATA	118	10	15	ESTs
GTTTCATTTT	119	10	15	ESTs
AATAAATTGC	120	10	15	ESTs
GTTTCATTTT	121	10	15	ESTs
ACACATTGTA	122	10	15	No match
TACCTATTGT	123	10	15	ESTs
TTTAGCAGAA	124	10	15	Homo sapiens cyclin E2 mRNA, complete cds
TTTAGCAGAA	125	10	15	ESTs
CAATTTATGA	126	9	13	ESTs
GTGAAGGTTT	127	9	13	Homo sapiens (huc) mRNA, complete cds
TGGACTTTTA	128	9	13	ESTs
CGATGCCACG	129	9	13	No match
GTGAAGGTTT	130	9	13	Neuron-specific RNA recognition motifs (RRMs)-containing protein [human, hippocampus, mRNA, 1992 nt]

Table 3. cont.

TGGACTTTTA	131	9	13	ESTs
CCTTCTTGTC	132	9	13	No match
TCCATTCAAG	133	9	13	Human clone 23586 mRNA sequence
CCTATGTATC	134	8	12	No match
ACGGACCAAT	135	8	12	No match
TATTATCTTG	136	8	12	ESTs
ACTTTATACG	137	8	12	ESTs
ACTTTATACG	138	8	12	ESTs, Weakly similar to EPIDERMAL GROWTH FACTOR RECEPTOR KINASE SUBSTRATE EPS8 (H.sapiens)
CGCAGTCCCC	139	8	12	BETA-NEOENDORPHIN-DYNORPHIN PRECURSOR
TGTAGTGCTC	140	8	12	No match
CTGCTTAAGT	141	8	12	ESTs, Weakly similar to unknown (H.sapiens)
ACAAGTGGAA	142	8	12	Human mRNA for KIAA0027 gene, partial cds
AATCCCAATG	143	7	10	Homo sapiens mRNA for KIAA0283 gene, partial cds
ACTATGCATC	144	7	10	No match
ACGAGTCATT	145	7	10	ESTs
TTACATTGTA	146	7	10	Homo sapiens clone 24461 mRNA sequence
ATGCCCCCTC	147	7	10	ESTs, Highly similar to HYPOTHETICAL 52.2 KD PROTEIN ZK512.6 IN CHROMOSOME III (Caenorhabditis elegans)
TTTTATTCAT	148	7	10	ESTs
ACAGAGCATT	149	7	10	No match
TGACCAATAG	150	7	10	No match
AATCCCAATG	151	7	10	Plastin 1 (I isoform)
Keratinocytes (0.087%)				
GCGAACTGGG	152	5	18	ORPHAN RECEPTOR TR4
GCAACACTAA	153	3	11	No match
GTAATGGATT	154	3	11	No match
AGCAGACGTG	155	3	11	No match
Breast Epithelium (0.14%)				
GGATTCGGTC	156	6	17	No match
CGGAAGGCGG	157	5	14	No match
TGTAAGTACG	158	5	14	No match
GATCAGTCAT	159	4	11	No match
GCTCAGAGTT	160	4	11	No match
Lung epithelium (0.17%)				
TAACCTCCCC	161	90	241	No match
AGGAACAACT	162	6	16	No match
GGGTCCGTGG	163	6	16	No match
TAGCAAAATA	164	5	13	No match
GCTGTGCACA	165	4	11	No match
CAGAAAATCA	166	4	11	No match

Table 3, cont.

GATTTGCTGG	167	4	11	No match
Melanocyte (0.93%)				
GTGCCATTCT	168	114	309	No match
GATATTTGTC	169	40	108	5,6-DIHYDROXYINDOLE-2-CARBOXYLIC ACID OXIDASE PRECURSOR
TATGATTTTA	170	39	106	ESTs
TCACTGCAAC	171	27	73	5,6-DIHYDROXYINDOLE-2-CARBOXYLIC ACID OXIDASE PRECURSOR
CCCAGTCACA	172	21	57	ESTs, Weakly similar to LACTOSE PERMEASE [Escherichia coli]
TATGAGAACC	173	17	46	ESTs, Highly similar to HIGH AFFIMMUNOGLOBULIN GAMMA FC RECEPTOR I PRECURSOR [Homo sapiens]
GAGTTTAGTG	174	16	43	No match
CTCCACTCTG	175	15	41	No match
ATCCAGTGAC	176	14	38	No match
TGATCTTGAG	177	14	38	ESTs, Moderately similar to PAS protein 5 [H.sapiens]
AATGGCTGTT	178	12	33	Human melanoma antigen recognized by T-cells (MART-1) mRNA
ATACTAAAAA	179	12	33	Human cysteine protease CPP32 isoform alpha mRNA, complete cds
ATACTAAAAA	180	12	33	EST
GTTTATTAAA	181	10	27	PROTEIN-TYROSINE PHOSPHATASE ZETA PRECURSOR
AGAAATCAGT	182	9	24	No match
TTGGATATTA	183	9	24	Homo sapiens clone 23785 mRNA sequence
AATTGAGTAG	184	9	24	Human DNA sequence from PAC 257A7 on chromosome 6p24. Contains two unknown genes and ESTs, STSs and a GSS
TGAGTGCTGC	185	9	24	No match
GCAGTACAGT	186	8	22	No match
GAATTCAGGA	187	7	19	Homo sapiens mRNA for KIAA0679 protein, partial cds
GACTTCTTTA	188	7	19	No match
GAATTCAGGA	189	7	19	Homo sapiens melastatin 1 (MLSN1) mRNA, complete cds
GTTTATACTG	190	7	19	No match
GAATTCAGGA	191	7	19	Homo sapiens mRNA for synaptosome associated protein of 23 kilodaltons, isoform A
GCCCGTGTAG	192	6	16	Msh (Drosophila) homeo box homolog 1 (formerly homeo box 7)
TGGGGTGTGC	193	6	16	Homo sapiens thyroid receptor interactor (TRIP8) mRNA, 3' end of cds
AATTTTTATG	194	5	14	Interferon regulatory factor 4
TCAGTGTCTG	195	5	14	ESTs
GGAGGTCAGC	196	5	14	ESTs
TTCTTCTCAA	197	5	14	ESTs
TTCTTCTCAA	198	5	14	ESTs
GGTTGTCTCT	199	5	14	ESTs, Weakly similar to line-1 protein ORF2 [H.sapiens]
CTTTGTTTAC	200	5	14	No match
CACTATAGAA	201	5	14	No match
TTTGGTTACA	202	4	11	EST
TCAAAACAAT	203	4	11	Human R kappa B mRNA, complete cds
TTTGGTTACA	204	4	11	Homo sapiens clone 23688 mRNA sequence
TATAGAGCAA	205	4	11	No match
TAATAACCAG	206	4	11	No match
TTCTATACTG	207	4	11	No match
GGAATACGGC	208	4	11	No match

Table 3, cont.

Prostate (0.05%)				
TGAACTGGCA	209	3	9	No match
AATGTTGGGG	210	3	9	No match
Normal Kidney (0.27%)				
CGACAAACTA	211	4	12	No match
GTAGCACAGA	212	4	12	No match
ACCGTCAATC	213	4	12	No match
TGGATCAGTC	214	4	12	Human mRNA for KIAA0259 gene, partial cds
TGGCTCGGTC	215	4	12	EST
GCGACTGCGA	216	4	12	No match
GCACTAGCTG	217	3	9	No match
GCGGCCGGTT	218	3	9	No match
CGGCAGTCCC	219	3	9	No match
GCCCACCTGT	220	3	9	No match
CGGCGGATGG	221	3	9	No match
CCCCAGGCCG	222	3	9	No match
CCCATTCCAA	223	3	9	No match
TCAAGAGGTG	224	3	9	No match
ATAACTGTTG	225	3	9	Human HFREP-1 mRNA for unknown protein, complete cds

Table 4. Ubiquitously expressed transcripts

Tag sequence	SEQ ID NO:	Copies/cell	Range	Range/Avg	Unigene Description
CATCTAAACT	266	44	22 - 62	0.91	Human mRNA for KIAA0038 gene, partial cds
GGGCAAGCCA	267	27	14 - 40	1.00	STEROID HORMONE RECEPTOR ERR1
ATTCAGCACC	268	29	11 - 40	1.03	ESTs, Highly similar to signal peptidase:SUBUNIT=12kD
TTGTTATTGC	269	15	6 - 21	1.04	Annexin VII (synexin)
ACAGGGTGAC	270	115	47 - 165	1.04	Homo sapiens mRNA for EDF-1 protein
GCTTCCATCT	271	39	17 - 58	1.06	H.sapiens BAT1 mRNA for nuclear RNA helicase (DEAD family)
GCTTCCATCT	272	39	17 - 58	1.06	BB1=malignant cell expression-enhanced gene/tumor progression-enhanced gene
GAGGGTGGCG	273	21	9 - 32	1.08	Human DR-nm23 mRNA, complete cds
GCAGGGTGGG	274	34	15 - 53	1.10	V-akt murine thymoma viral oncogene homolog 2
AGCCCTCCCT	275	85	42 - 136	1.12	Homo sapiens autoantigen p542 mRNA, complete cds
ATGGCCATAG	276	15	5 - 22	1.12	Human mRNA for YSK1, complete cds
GTGGGTGTCC	277	20	9 - 32	1.13	ESTs
TGTAGTTTGA	278	41	14 - 62	1.14	Transcription elongation factor B (SIII), polypeptide 1-like
GGGGCTGTGG	279	14	6 - 21	1.15	Human TFIIIC Box B-binding subunit mRNA, complete cds
GGGGCTGTGG	280	14	6 - 21	1.15	Homo sapiens mRNA for smallest subunit of ubiquinol-cytochrome c reductase, complete cds
CACGCAATGC	281	111	53 - 182	1.17	Human homolog of Drosophila enhancer of split m9/m10 mRNA, complete cds
CTCACACATT	282	49	20 - 78	1.18	LYSOSOME-ASSOCIATED MEMBRANE GLYCOPROTEIN 1 PRECURSOR
CAAATGAGGA	283	36	15 - 58	1.19	Neuroblastoma RAS viral (v-ras) oncogene homolog
TGTAAGTCTG	284	21	8 - 33	1.19	Human p62 mRNA, complete cds
ACCAAGGAGG	285	63	25 - 100	1.19	ESTs
ACCAAGGAGG	286	63	25 - 100	1.19	DNA-DIRECTED RNA POLYMERASE II 23 KD POLYPEPTIDE
ACCAAGGAGG	287	63	25 - 100	1.19	Human mRNA for transcription elongation factor S-II, hS-II-T1, complete cds
TGAGGCAGGG	288	17	7 - 27	1.20	Syntaxin 5A
TCCACGCACC	289	39	14 - 61	1.20	ESTs
TAGGGCAATC	290	40	14 - 62	1.21	H.sapiens mRNA for SMT3B protein
GGTAGCCTGG	291	61	25 - 98	1.21	Damage-specific DNA binding protein 1 (127 kD)
TCAACAGCCA	292	14	6 - 23	1.21	Human translation initiation factor 3 47 kDa subunit mRNA, complete cds
CTCTGTGTGG	293	18	7 - 29	1.21	Homo sapiens EB1 mRNA, complete cds
CCTATTTACT	294	115	51 - 193	1.23	Cytochrome c oxidase subunit IV
TGCATCTGGT	295	104	32 - 162	1.24	78 KD GLUCOSE REGULATED PROTEIN PRECURSOR
GCTCTCTATG	296	72	21 - 111	1.25	H.sapiens mRNA for rat translocon-associated protein delta homolog
GAAGGCATCC	297	39	16 - 64	1.25	PROBABLE 26S PROTEASE SUBUNIT TBP-1
CCACTCCTCA	298	59	19 - 93	1.26	DEFENDER AGAINST CELL DEATH 1
GCTGTCATCA	299	31	8 - 47	1.27	26S PROTEASE REGULATORY SUBUNIT 4
CGGCTGGTGA	300	63	24 - 105	1.28	Proteasome component C5
AAGCCAGGAC	301	65	26 - 110	1.31	Homo sapiens chromosome 19, cosmid R32469
TGAGAGGGTG	302	32	15 - 57	1.32	14-3-3 PROTEIN TAU
GCGTGATCCT	303	33	10 - 54	1.32	ALCOHOL DEHYDROGENASE
CTGCCAACTT	304	51	11 - 78	1.33	COFILIN, NON-MUSCLE ISOFORM
CCAAACGTGT	305	148	56 - 254	1.33	HISTONE H3.3
GCGGGAGGGC	306	45	12 - 72	1.34	ADP-RIBOSYLATION FACTOR-LIKE PROTEIN 2
GGCCAGCCCT	307	70	20 - 114	1.34	ESTs

Table 4, cont.

GGCCAGCCCT	308	70	20	114	1.34	Phosphofructokinase (liver type)
TGGGCAAAGC	309	608	189	1014	1.36	Translation elongation factor 1 gamma
GCAAAACCAG	310	29	12	52	1.36	Human mRNA for KIAA0002 gene, complete cds
ACTTACCTGC	311	107	33	179	1.36	Cytochrome c oxidase subunit VIb
GTTGGTCTGT	312	32	11	54	1.36	ESTs
TGCTACTGGT	313	18	7	32	1.36	Surfeit 1
GACGACACGA	314	401	71	618	1.37	Ribosomal protein S28
CAAGTGGCAA	315	18	5	31	1.37	Homo sapiens Grf40 adaptor protein (Grf40) mRNA, complete cds
TACTCTTGGC	316	72	16	114	1.37	HETEROGENEOUS NUCLEAR RIBONUCLEOPROTEIN L
GACTGTGCCA	317	75	15	118	1.37	Human cytoplasmic dynein light chain 1 (hdlc1) mRNA, complete cds
TTGCCGGTTA	318	19	9	34	1:37	Homo sapiens clone 24592 mRNA sequence
CATTGCAGGA	319	14	5	25	1.38	Homo sapiens Chromosome 16 BAC clone CIT987SK-A-152E5
CAGGAACGGG	320	97	26	159	1.38	DUAL SPECIFICITY MITOGEN-ACTIVATED PROTEIN KINASE KINASE 2
AATAGGTCCA	321	219	64	371	1.40	Ribosomal protein S25
ACCTCAGGAA	322	67	32	126	1.41	Human high density lipoprotein binding protein (HBP) mRNA, complete cds
ATGACTCAAG	323	26	12	48	1.41	Human mRNA for protein tyrosine phosphatase (PTP-BAS, type 2), complete cds
ATGACTCAAG	324	26	12	48	1.41	Homo sapiens mRNA, chromosome 1 specific transcript KIAA0488
GCCTCTGCCA	325	26	12	46	1.41	Human mRNA for KIAA0272 gene, partial cds
TGCTTGTCCC	326	62	25	112	1.42	ADP-ribosylation factor 1
GGTGGCACTC	327	112	41	199	1.42	Aplysia ras-related homolog 12
GGGCTGGGGT	328	659	168	1102	1.42	H.sapiens mRNA for ribosomal protein L29
GGGCTGGGGT	329	659	168	1102	1.42	Homo sapiens sperm acrosomal protein mRNA, complete cds
CACAAACGGT	330	844	252	1449	1.42	40S RIBOSOMAL PROTEIN S27
CATTGAAGGG	331	37	13	66	1.42	Homo sapiens clone 24433 myelodysplasia/myeloid leukemia factor 2 mRNA, complete cds
GTGACTGCCA	332	38	15	69	1.42	DPH2L=candidate tumor suppressor gene (ovarian cancer critical region of deletion)
GTGACTGCCA	333	38	15	69	1.42	Homo sapiens clone 24722 unknown mRNA, partial cds
AAGACAGTGG	334	678	222	1190	1.43	Ribosomal protein L37a
CTGGCTGCAA	335	86	24	147	1.43	Cytochrome c oxidase subunit Vb
ACCGGGAGGT	336	18	5	30	1.43	Human DNA from chromosome 19-specific cosmid R27090, genomic sequence
ATGGAGACTT	337	26	8	46	1.43	Homo sapiens citrate synthase mRNA, complete cds
CAGCTCATCT	338	40	17	74	1.44	Homo sapiens hJTB mRNA, complete cds
ACGTGGTGAT	339	52	6	81	1.44	ESTs, Highly similar to LEYDIG CELL TUMOR 10 KD PROTEIN [Rattus norvegicus]
GCGGTGAGGT	340	37	9	62	1.44	Homo sapiens small glutamine-rich tetratricopeptide repeat (TPR) containing protein
GTGGCACACG	341	105	24	176	1.44	Eukaryotic translation initiation factor 3 (eIF-3) p36 subunit
GTGACAACAC	342	42	11	71	1.45	Voltage-dependent anion channel 1
CTGCTATACG	343	226	70	396	1.45	Ribosomal protein L5
ACTGGCTGCT	344	27	10	50	1.46	ESTs
GGAAGCACGG	345	53	16	93	1.46	Human antisecretory factor-1 mRNA, complete cds
GGAAGCACGG	346	53	16	93	1.46	Tag matches ribosomal RNA sequence
CTGTTGGTGA	347	295	86	516	1.46	40S RIBOSOMAL PROTEIN S23
TCAGATCTTT	348	358	141	663	1.46	Ribosomal protein S4, X-linked
TGGAATGCTG	349	78	37	151	1.46	Homo sapiens NADH:ubiquinone dehydrogenase 51 kDa subunit (NDUFV1) mRNA, nuclear gene encoding mitochondrial protein, complete cds
TAAGGAGCTG	350	289	71	493	1.46	Ribosomal protein S26
GGCTTTGGAG	351	41	15	75	1.46	ESTs

Table 4, cont.

CGCACCATTG	352	41	14	74	1.46	GCN5-like 1=GCN5 homolog/putative regulator of transcriptional activation (clone GCN5L1)
CGCTGGTTCC	353	443	177	825	1.46	Homo sapiens ribosomal protein L11 mRNA, complete cds
GGGCCTGGGG	354	62	13	105	1.46	ESTs
CTCGAGGAGG	355	43	10	73	1.47	Human ribosomal protein L23-related mRNA, complete cds
TTGGTCCTCT	356	1233	363	2177	1.47	60S RIBOSOMAL PROTEIN L41
TCCCTGGCAT	357	15	5	27	1.47	Heterogeneous nuclear ribonucleoprotein K
GGGGGCTGCT	358	11	6	23	1.47	ESTs
GGGGGCTGCT	359	11	6	23	1.47	Human lysyl oxidase-related protein (WS9-14) mRNA, complete cds
CCACCCCGAA	360	109	14	174	1.48	Testis enhanced gene transcript
CTGCTAGGAA	361	21	9	40	1.48	H.sapiens mRNA for TRAMP protein
AACTGCGGCA	362	15	7	29	1.48	ESTs
TGGAGTGGAG	363	134	56	254	1.48	Human guanylate kinase (GUK1) mRNA, complete cds
TGAAGGAGCC	364	107	33	191	1.48	ATP SYNTHASE LIPID-BINDING PROTEIN P2 PRECURSOR
GGGGACTGAA	365	77	24	138	1.48	Homo sapiens mRNA for low molecular mass ubiquinone-binding protein, complete cds
TGCACGTTTT	366	526	196	979	1.49	Human mRNA for antileukoprotease (ALP) from cervix uterus
CTGGATGCCG	367	33	11	59	1.49	Radin blood group
CCCCCTCGTG	368	24	8	44	1.49	Adrenergic, beta, receptor kinase 1
ATGATGCGGT	369	41	13	74	1.49	Cytoplasmic antiproteinase=38 kda intracellular serine proteinase inhibitor
ATTCTCCAGT	370	356	86	618	1.50	Ribosomal protein L17
CCCCAGTTGC	371	219	90	418	1.50	Calpain, small polypeptide
CCAAGGATTG	372	21	6	38	1.50	Solute carrier family 5 (sodium/glucose cotransporter), member 2
GACCGAGGTG	373	25	6	43	1.50	Ewing sarcoma breakpoint region 1
GACTCTCTCA	374	13	5	25	1.50	ESTs
GACTCTGGGA	375	21	6	37	1.51	ESTs, Moderately similar to T13H5.2 [C.elegans]
GACTCTGGGA	376	21	6	37	1.51	Actin, gamma 1
CGCCGCGGTG	377	207	54	368	1.51	Homo sapiens Chromosome 16 BAC clone CIT987SK-A-761H5
CCAGAACAGA	378	361	119	666	1.52	60S RIBOSOMAL PROTEIN L30
CCAGAACAGA	379	361	119	666	1.52	Deoxythymidylate kinase
TGGTTTTTGG	380	26	5	43	1.52	Homo sapiens acyl-protein thioesterase mRNA, complete cds
TTTTTGTACA	381	38	13	71	1.52	ER LUMEN PROTEIN RETAINING RECEPTOR 1
GTTCTCCCAC	382	65	24	122	1.52	ESTs, Highly similar to PROTEIN TRANSPORT PROTEIN SEC61 ALPHA SUBUNIT
GACCCTGCCC	383	192	30	323	1.52	Human FK-506 binding protein homologue (FKBP38) mRNA, complete cds
GCCCGCCTTG	384	49	16	91	1.52	Homo sapiens (clone mf.18) RNA polymerase II mRNA, complete cds
GGTGCTGGAG	385	24	8	45	1.53	Homo sapiens mRNA for putative methyltransferase
TTACCTCCTT	386	78	21	141	1.53	Homo sapiens 3-phosphoglycerate dehydrogenase mRNA, complete cds
AAACCAGGGC	387	18	5	33	1.53	ESTs
TTCTGGCTGC	388	85	11	141	1.53	Ubiquinol-cytochrome c reductase core protein I
TTCTGGCTGC	389	85	11	141	1.53	Human BAC clone RG114A06 from 7q31
CTTCTCACCG	390	33	8	58	1.54	Ubiquitin-conjugating enzyme E2I (homologous to yeast UBC9)
GAGAACCGTA	391	48	13	87	1.54	ESTs, Moderately similar to regulatory protein
GCGACCGTCA	392	658	51	1076	1.56	Aldolase A
GTCAAGACCA	393	28	11	54	1.56	Adaptin, beta 1 (beta prime)
CTGGGTCTCC	394	42	12	78	1.56	60S RIBOSOMAL PROTEIN L13
CGATTCTGGA	395	27	11	53	1.56	H.sapiens mRNA for ras-related GTP-binding protein
CAGGAGGAGT	396	73	19	132	1.56	PROBABLE PROTEIN DISULFIDE ISOMERASE ER-60 PRECURSOR
CAAAATCAGG	397	44	12	81	1.56	Human mRNA for cyclin I, complete cds

Table 4, cont.

CTGGGTTAAT	398	615	116	-	1081	1.57	40S RIBOSOMAL PROTEIN S19
TTTTCTGCTG	399	34	6	-	60	1.57	Hydroxyacyl-Coenzyme A dehydrogenase/3-ketoacyl-Coenzyme A thiolase/enoyl-Coenzyme A hydratase (trifunctional protein), beta subunit
CCCTGGCAAT	400	30	14	-	81	1.57	ESTs
AGGCTACGGA	401	807	199	-	1472	1.58	60S RIBOSOMAL PROTEIN L13A
GAGGCCATCC	402	23	8	-	45	1.58	Homo sapiens chromosome 19, cosmid R30783
CTTTGATGTT	403	26	11	-	52	1.58	Homo sapiens mRNA for NORI-1, complete cds
TTGGACCTGG	404	113	29	-	206	1.58	ESTs, Weakly similar to MALONYL COA-ACYL CARRIER PROTEIN TRANSACYLASE [E.coli]
TTGGACCTGG	405	113	29	-	206	1.58	ATP synthase, H+ transporting, mitochondrial F1 complex, delta subunit
GTTCGTGCCA	406	213	43	-	379	1.58	Ribosomal protein L35a
GATGCTGCCA	407	154	34	-	277	1.58	Human mRNA for Epstein-Barr virus small RNAs (EBERs)associated protein (EAP)
ACGGCTCCGA	408	27	8	-	50	1.58	ESTs
GAGTCAGGAG	409	29	6	-	53	1.59	ESTs, Highly similar to COATOMER ZETA SUBUNIT [Bos taurus]
GGAGGCTGAG	410	84	37	-	171	1.59	Homo sapiens mRNA for KIAA0792 protein, complete cds
GGAGGCTGAG	411	84	37	-	171	1.59	Homo sapiens putative fatty acid desaturase MLD mRNA, complete cds
GTGATGGTGT	412	75	24	-	143	1.59	Thyroid autoantigen 70kD (Ku antigen)
TCAGATGGCG	413	45	6	-	78	1.59	Homo sapiens hD54+ins2 isoform (hD54) mRNA, complete cds
ATGCGAAAGG	414	32	9	-	59	1.59	Dodecenoyl-Coenzyme A delta isomerase (3,2 trans-enoyl-Coenzyme A isomerase)
TGCTGGGTGG	415	67	26	-	133	1.60	ESTs, Highly similar to NADH-UBIQUINONE OXIDOREDUCTASE ASHI SUBUNIT PRECURSOR [Bos taurus]
TGCTGGGTGG	416	67	26	-	133	1.60	Homo sapiens folylpolyglutamate synthetase mRNA, complete cds
TCAAATGCAT	417	37	9	-	68	1.60	HETEROGENEOUS NUCLEAR RIBONUCLEOPROTEINS C1/C2
TCCAAGGAAG	418	13	5	-	26	1.60	Homo sapiens DBI-related protein mRNA, complete cds
CCCAGGGAGA	419	49	11	-	90	1.60	Homo sapiens chaperonin containing t-complex polypeptide 1, delta subunit (Cctd) mRNA, complete cds
TGGCCTGCCC	420	54	15	-	102	1.60	ESTs
TGGCCTGCCC	421	54	15	-	102	1.60	ESTs, Moderately similar to PEANUT PROTEIN [Drosophila melanogaster]
GGCCAAAGGC	422	39	14	-	77	1.60	Human mRNA for KIAA0064 gene, complete cds
GGCCTGCTGC	423	69	13	-	125	1.60	ESTs, Highly similar to C10 [H.sapiens]
GTGAAGCTGA	424	22	7	-	41	1.61	ESTs, Highly similar to HYPOTHETICAL 6.3 KD PROTEIN ZK652.2 IN CHROMOSOME III [Caenorhabditis elegans]
GTGAAGCTGA	425	22	7	-	41	1.61	ESTs, Highly similar to thymic epithelial cell surface antigen [M.musculus]
GAAATGTAAG	426	50	12	-	93	1.62	ESTs
GAAATGTAAG	427	50	12	-	93	1.62	H.sapiens hnRNP-E2 mRNA
CGTGTTAATG	428	73	31	-	148	1.62	CELLULAR NUCLEIC ACID BINDING PROTEIN
AGGGGATTCC	429	19	9	-	40	1.62	Human arginine-rich protein (ARP) gene, complete cds
CAGCTCACTG	430	166	23	-	326	1.63	Homo sapiens CAG-isl 7 mRNA, complete cds
GTTTGGCAGT	431	35	13	-	70	1.63	Homo sapiens mRNA for EDF-1 protein
GGAGCTCTGT	432	48	13	-	92	1.63	ESTs, Moderately similar to NADH-UBIQUINONE OXIDOREDUCTASE B15 SUBUNIT [Bos taurus]
TGGAACTGTG	433	22	5	-	42	1.63	ESTs, Weakly similar to !!!! ALU SUBFAMILY SQ WARNING ENTRY !!!! [H.sapiens]
TCTGCTTACA	434	58	18	-	114	1.63	Human ribosomal protein L10 mRNA, complete cds
AGGGCTTCCA	435	643	205	-	1257	1.64	UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX SUBUNIT VI REQUIRING PROTEIN
GAGCAAACGG	436	20	5	-	37	1.64	Homo sapiens chromosome 19, cosmid R26445
TGTGATCAGA	437	88	27	-	171	1.64	Homo sapiens F1F0-type ATP synthase subunit g mRNA, complete cds

Table 4, cont.

ACACTACGGG	438	37	6	66	1.64	ESTs, Weakly similar to putative progesterone binding protein [H.sapiens]
AGCCAAAAAA	439	41	12	79	1.64	H.sapiens hnRNP-E2 mRNA
GCGGGTGTGG	440	16	5	32	1.64	Human methionine aminopeptidase mRNA, complete cds
TTGCTAGAGG	441	39	13	78	1.65	ESTs, Weakly similar to F35H10.6 gene product [C.elegans]
GGGGCTTCTG	442	15	6	30	1.65	Human mRNA for cysteine protease, complete cds
AACTCTTGAA	443	45	14	87	1.65	Human translation initiation factor eIF3 p40 subunit mRNA, complete cds
GTCTGACCCC	444	44	8	80	1.65	PROTEIN PHOSPHATASE PP2A, 65 KD REGULATORY SUBUNIT, ALPHA ISOFORM
ATGTCATCAA	445	48	12	92	1.65	Human clathrin assembly protein 50 (AP50) mRNA, complete cds
TCTGTCAAGA	446	40	15	81	1.66	ATP synthase, H+ transporting, mitochondrial F1 complex, O subunit (oligomycin sensitivity conferring protein)
GCCCCAGCGA	447	23	8	46	1.66	ESTs
GGCAAGCCCC	448	425	119	824	1.66	Heat shock 27kD protein 1
CTCATCAGCT	449	48	16	95	1.66	ADENYLYL CYCLASE-ASSOCIATED PROTEIN 1
CTGTTGATTG	450	137	49	276	1.66	Heterogeneous nuclear ribonucleoprotein A1
GCTTTTAAGG	451	171	27	312	1.66	40S RIBOSOMAL PROTEIN S20
GCCTGAGCCT	452	13	6	28	1.66	ESTs
GAGCGGGATG	453	57	21	116	1.66	Proteasome (prosome, macropain) subunit, beta type, 6
TTCACAGTGG	454	56	13	107	1.67	Calcineurin B
GCCCGTGCCA	455	23	8	46	1.67	ESTs, Highly similar to HYPOTHETICAL 38.2 KD PROTEIN IN BEM2-SPT2 INTERGENIC REGION [Saccharomyces cerevisiae]
CCCTAGGTTG	456	51	14	98	1.67	Human mRNA for KIAA0315 gene, partial cds
CCCTGATTTT	457	33	12	66	1.67	Human p97 mRNA, complete cds
GTGTTAACCA	458	314	73	599	1.67	Human ribosomal protein L10 mRNA, complete cds
AGGAAAGCTG	459	469	162	948	1.68	ESTs, Highly similar to 60S RIBOSOMAL PROTEIN L36 [Rattus norvegicus]
TTCTCTCTGT	460	31	8	60	1.68	ADP-ribosylation factor 5
TTACTAAATG	461	26	5	48	1.68	Calnexin
GGGTGTGGTG	462	18	5	36	1.68	ESTs
CCACTGCAGT	463	14	5	29	1.68	GLYCOPROTEIN HORMONES ALPHA CHAIN PRECURSOR
AGCCTGGACT	464	47	17	95	1.69	Human mRNA for Mr 110,000 antigen, complete cds
GTGGGGTGAC	465	24	8	47	1.69	ESTs, Weakly similar to HYPOTHETICAL 21.5 KD PROTEIN IN SEC15-SAP4 INTERGENIC REGION [S.cerevisiae]
CACTACACGG	466	46	11	88	1.69	FK506-BINDING PROTEIN PRECURSOR
CTCATAGCAG	467	92	31	187	1.69	TRANSLATIONALLY CONTROLLED TUMOR PROTEIN
GGAATGTACG	468	94	27	187	1.70	Human mitochondrial ATP synthase subunit 9, P3 gene copy, mRNA, nuclear gene encoding mitochondrial protein, complete cds
CTGAGGGTGG	469	17	8	38	1.70	ESTs
AAGGTCGAGC	470	75	9	136	1.70	60S RIBOSOMAL PROTEIN L24
GAATCACTGC	471	18	5	35	1.70	Homo sapiens ribosomal protein L33-like protein mRNA, complete cds
ACATCATCGA	472	374	86	722	1.70	Ribosomal protein L12
GAATGAGGAC	473	27	6	51	1.70	Human mRNA for reticulocalbin, complete cds
CCTCGCTCAG	474	44	14	89	1.70	Hydroxyacyl-Coenzyme A dehydrogenase/3-ketoacyl-Coenzyme A thiolase/enoyl-Coenzyme A hydratase (trifunctional protein), alpha subunit
TCCTAGCCTG	475	16	5	33	1.70	Homo sapiens SPF31 (SPF31) mRNA, complete cds
AGGTGCGGGG	476	35	5	64	1.71	Human hASNA-I mRNA, complete cds
CTCCAATAAA	477	14	7	31	1.71	Homo sapiens clone 24775 mRNA sequence

Table 4, cont.

GCGCTGGAGT	478	73	23	-	147	1.71	ESTs, Weakly similar to HYPOTHETICAL 9.9 KD PROTEIN B0495.6 IN CHROMOSOME II [C.elegans]
AATTTGCAAC	479	21	5	-	40	1.71	Homo sapiens histone macroH2A1.2 mRNA, complete cds
AACGCGGCCA	480	448	22	-	790	1.71	Macrophage migration inhibitory factor
GGTGTATATG	481	21	7	-	42	1.71	Homo sapiens chromosome 9, P1 clone 11659
GGCAACAAAA	482	35	6	-	66	1.71	Human (clone E5.1) RNA-binding protein mRNA, complete cds
GGCAACAAAA	483	35	6	-	66	1.71	Homo sapiens importin beta subunit mRNA, complete cds
TTTGTGACTG	484	28	13	-	62	1.71	Homo sapiens phosphoprotein CtBP mRNA, complete cds
ATGAGGCCGG	485	23	7	-	47	1.72	No match
TCAGTTTGTC	486	39	15	-	81	1.72	Human HS1 binding protein HAX-1 mRNA, nuclear gene encoding mitochondrial protein, complete cds
CCCTATTAAG	487	69	10	-	129	1.72	No match
TTTCTAGTTT	488	55	28	-	123	1.72	Human mRNA for KIAA0108 gene, complete cds
GGGCCCTTCC	489	20	5	-	40	1.72	Homo sapiens clone 24684 mRNA sequence
GGGCCCTTCC	490	20	5	-	40	1.72	Fibulin 1
CCTTGGTTTT	491	24	6	-	47	1.72	Homo sapiens DNA-binding protein (CROC-1B) mRNA, complete cds
GCTAAGGAGA	492	61	21	-	161	1.72	Human ras-related C3 botulinum toxin substrate (rac) mRNA, complete cds
TGAGGGGTGA	493	27	8	-	56	1.72	Human Gps1 (GPS1) mRNA, complete cds
CCAGCTGCCA	494	63	19	-	126	1.73	Ubiquitin activating enzyme E1
GGGCTGTTTG	495	16	5	-	34	1.73	No match
TGGACACAAG	496	18	5	-	36	1.73	Arginyl-tRNA synthetase
TCTCCAGGAA	497	44	12	-	89	1.73	ESTs, Weakly similar to PUTATIVE MITOCHONDRIAL CARRIER C16C10.1 [C.elegans]
TGATGTTTGA	498	24	8	-	49	1.73	Human mRNA for KIAA0058 gene, complete cds
GTGGTGCACG	499	62	13	-	155	1.73	No match
GTCTGCACCT	500	32	8	-	64	1.73	ESTs, Weakly similar to NUCLEAR PROTEIN SNF7 [Saccharomyces cerevisiae]
GATGACCCCG	501	32	11	-	66	1.73	ESTs, Weakly similar to F08G12.1 [C.elegans]
ATCAAGGGTG	502	269	27	-	494	1.73	Ribosomal protein L9
TCTGGTCTGG	503	34	12	-	72	1.74	Human surface antigen mRNA, complete cds
AGGATGACCC	504	42	6	-	79	1.74	ESTs, Weakly similar to Ion channel homolog RIC [M.musculus]
AAAGGGGGCA	505	28	9	-	58	1.74	H.sapiens mRNA for activin beta-C chain
GGCTTTACCC	506	178	56	-	365	1.74	Eukaryotic translation initiation factor 5A
GCTTTTTAGA	507	39	10	-	78	1.74	Human non-histone chromosomal protein HMG-14 mRNA, complete cds
CTCTGCTCGG	508	18	6	-	37	1.74	Homo sapiens clone 638 unknown mRNA, complete sequence
GCCTGGGACT	509	58	28	-	130	1.74	ESTs
GGTAGCAGGG	510	26	5	-	50	1.74	Homo sapiens clone 23930 mRNA sequence
GCCGATCCTC	511	31	7	-	61	1.74	Homo sapiens cofactor A protein mRNA, complete cds
GCAGCTCAGG	512	50	13	-	101	1.74	Cathepsin D (lysosomal aspartyl protease)
CGCAGTGTCC	513	118	20	-	225	1.75	Vacuolar H+ ATPase proton channel subunit
CCCCTATTAA	514	62	13	-	121	1.75	No match
TTGTAAAAGG	515	23	8	-	47	1.75	Homo sapiens chromosome 9, P1 clone 11659
CCACACCGGT	516	17	6	-	36	1.75	Heme oxygenase (decycling) 2
CCTGGAAGAG	517	192	60	-	396	1.75	Procollagen-proline, 2-oxoglutarate 4-dioxygenase (proline 4-hydroxylase), beta polypeptide (protein disulfide isomerase; thyroid hormone binding protein p55)
TAGCCGCTGA	518	37	7	-	72	1.75	Homo sapiens alpha SNAP mRNA, complete cds
CCTAGGACCT	519	19	5	-	39	1.75	Homo sapiens Arp2/3 protein complex subunit p20-Arc (ARC20) mRNA, complete cds
GTGGACCCTG	520	26	9	-	54	1.75	Surfeit 1

Table 4, cont.

GTGGACCCTG	521	26	9	-	54	1.75	ESTs, Weakly similar to R05G6.4 gene product [C.elegans]
TTGGGAGCAG	522	32	6	-	63	1.76	Isoleucine-tRNA synthetase
GTCTCACGTG	523	23	9	-	49	1.76	ESTs
GTACTGTGGC	524	114	24	-	225	1.76	Homo sapiens nuclear chloride ion channel protein (NCC27) mRNA, complete cds
AAGATAATGC	525	12	5	-	27	1.76	ESTs, Weakly similar to Yel007c-ap [S.cerevisiae]
AATACCTCGT	526	31	7	-	61	1.76	ESTs
ACCTTGTGCC	527	23	6	-	47	1.76	ESTs, Weakly similar to alpha 2,6-sialyltransferase [R.norvegicus]
ACCTTGTGCC	528	23	6	-	47	1.76	Sorbitol dehydrogenase
GGAGGGGGCT	529	88	16	-	172	1.77	LAMIN A
GCCTATGGTC	530	39	9	-	78	1.77	ESTs, Highly similar to SEX-REGULATED PROTEIN JANUS-A [Drosophila melanogaster]
GTGCTGAATG	531	459	219	-	1031	1.77	MYOSIN LIGHT CHAIN ALKALI, SMOOTH-MUSCLE ISOFORM
TCGTCGCAGA	532	37	9	-	75	1.77	ESTs, Highly similar to NADH-UBIQUINONE OXIDOREDUCTASE SUBUNIT B14.5A [Bos taurus]
GTGACAGAAG	533	178	36	-	351	1.77	Eukaryotic translation initiation factor 4A (eIF-4A) isoform 1
TCAACGGTGT	534	15	5	-	31	1.77	Homo sapiens mRNA for RanBPM, complete cds
GAGCCTTGGT	535	58	11	-	113	1.77	Protein phosphatase 1, catalytic subunit, alpha isoform
TACATCCGAA	536	19	6	-	40	1.78	ESTs
GTCTGTGAGA	537	29	12	-	64	1.78	Homo sapiens mRNA for Hrs, complete cds
GTTAACGTCC	538	95	18	-	187	1.78	Homo sapiens Bruton's tyrosine kinase (BTK), alpha-D-galactosidase A (GLA), L44-like ribosomal protein (L44L) and FTP3 (FTP3) genes, complete cds
GTGCGCTAGG	539	141	27	-	277	1.78	ESTs, Weakly similar to F49C12.12 [C.elegans]
CGGATAAGGC	540	17	6	-	36	1.78	ESTs
GTCTGGGGCT	541	204	49	-	413	1.78	SM22-ALPHA HOMOLOG
CATCCTGCTG	542	64	12	-	125	1.78	Human mRNA for 26S proteasome subunit p97, complete cds
TCACAAGCAA	543	142	52	-	305	1.78	H.sapiens alpha NAC mRNA
GGCTGATGTG	544	73	15	-	146	1.78	Glycyl-tRNA synthetase
CCCGTCCGGA	545	1272	293	-	2564	1.78	60S RIBOSOMAL PROTEIN L13
TCCGCGAGAA	546	98	33	-	208	1.78	ESTs, Weakly similar to SEX-DETERMINING TRANSFORMER PROTEIN 1 [Caenorhabditis elegans]
GTGCTGGAGA	547	98	12	-	187	1.79	Human SnRNP core protein Sm D2 mRNA, complete cds
TCCTCAAGAT	548	26	8	-	54	1.79	Human enhancer of rudimentary homolog mRNA, complete cds
CAACTTAGTT	549	60	20	-	127	1.79	Human myosin regulatory light chain mRNA, complete cds
GGGCAGCTGG	550	35	12	-	75	1.79	ESTs
TTTCAGAGAG	551	43	8	-	84	1.79	Human calmodulin mRNA, complete cds
TTTCAGAGAG	552	43	8	-	84	1.79	Signal recognition particle 9 kD protein
GACGCAGAAG	553	17	6	-	36	1.79	ESTs, Highly similar to ALPHA-ADAPTIN [Mus musculus]
GGAAGTTTCG	554	35	9	-	72	1.79	ESTs, Weakly similar to similar to oxysterol-binding proteins: partial CDS [C.elegans]
GTTGCTGCCC	555	34	5	-	65	1.79	Homo sapiens mRNA for putative seven transmembrane domain protein
GCTGGGGTGG	556	21	6	-	44	1.79	H.sapiens mRNA for mediator of receptor-induced toxicity
CTCAACATCT	557	456	99	-	918	1.80	Ribosomal protein, large, P0
CAAGCAGGAC	558	42	8	-	84	1.80	ESTs, Weakly similar to transmembrane protein [H.sapiens]
TTGGCTTTTC	559	27	8	-	57	1.80	ESTs
TGGCAACCTT	560	38	17	-	85	1.80	ESTs, Highly similar to GLUTATHIONE S-TRANSFERASE, MITOCHONDRIAL [Rattus norvegicus]
GCATAATAGG	561	391	83	-	786	1.80	Ribosomal protein L21
GGGGGTAACT	562	43	9	-	86	1.80	RNA-BINDING PROTEIN FUS/TLS

Table 4, cont.

CCTTCGAGAT	563	274	55	-	549	1.80	Ribosomal protein S5
CGGGCCGTGC	564	18	6	-	38	1.80	H.sapiens mRNA for Glyoxalase II
GTGTTGCACA	565	210	42	-	421	1.80	Ribosomal protein S13
CCTCGGAAAA	566	158	27	-	312	1.81	60S RIBOSOMAL PROTEIN L38
AATAAAGGCT	567	56	9	-	110	1.81	Myosin, light polypeptide 3, alkali; ventricular, skeletal, slow
AATAAAGGCT	568	56	9	-	110	1.81	Aplysia ras-related homolog 9
CTTCTGTGTA	569	21	9	-	47	1.81	Homo sapiens immunophilin homolog ARA9 mRNA, complete cds
CTTCTGTGTA	570	21	9	-	47	1.81	Human mRNA for KIAA0190 gene, partial cds
GGTCCAGTGT	571	144	28	-	286	1.81	Phosphoglycerate mutase 1 (brain)
AGCACCTCCA	572	701	197	-	1467	1.81	Eukaryotic translation elongation factor 2
AAGCTGAGTG	573	39	12	-	82	1.81	Human M4 protein mRNA, complete cds
GTTTCTTCCC	574	27	11	-	60	1.81	ESTs
TGAGGGAATA	575	191	51	-	397	1.82	Triosephosphate isomerase 1
AGCTCTCCCT	576	447	150	-	962	1.82	60S RIBOSOMAL PROTEIN L23
TACGTTGCAG	577	18	8	-	40	1.82	Homo sapiens GC20 protein mRNA, complete cds
GGGTGTGTAT	578	16	6	-	35	1.82	Homo sapiens angio-associated migratory cell protein (AAMP) mRNA, complete cds
GGAGGGATCA	579	37	12	-	79	1.82	Homo sapiens integrin-linked kinase (ILK) mRNA, complete cds
ATCAGTGGCT	580	64	25	-	143	1.82	PROTEASOME BETA CHAIN PRECURSOR
CCCCCTGCCC	581	57	17	-	121	1.83	ESTs
CCCCCTGCCC	582	57	17	-	121	1.83	ESTs
CAAAAAAAAA	583	94	8	-	180	1.83	Cholinergic receptor, nicotinic, alpha polypeptide 3
ACCTGCCGAC	584	18	5	-	37	1.83	Homo sapiens growth suppressor related (DOC-1R) mRNA, complete cds
GACCAGAAAA	585	81	17	-	165	1.83	CYTOCHROME C OXIDASE POLYPEPTIDE VIA-LIVER PRECURSOR
AGCCACTGCG	586	33	9	-	69	1.83	No match
TTGAGCCAGC	587	43	21	-	101	1.83	Human KH type splicing regulatory protein KSRP mRNA, complete cds
TTTCAGGGGA	588	51	9	-	103	1.84	ESTs, Moderately similar to N-methyl-D-aspartate receptor glutamate-binding chain [R.norvegicus]
TCCGGCCGCG	589	75	32	-	169	1.84	ESTs
GTGATCTCCG	590	22	6	-	46	1.84	ESTs
CTGCTGAGTG	591	46	6	-	90	1.84	ESTs, Highly similar to HYPOTHETICAL 14.1 KD PROTEIN C31A2.02 IN CHROMOSOME I [Schizosaccharomyces pombe]
CTGCTTAAGG	592	16	6	-	36	1.84	ESTs, Highly similar to HYPOTHETICAL 68.7 KD PROTEIN ZK757.1 IN CHROMOSOME III [Caenorhabditis elegans]
TGTGGCCTCC	593	33	14	-	74	1.84	ESTs, Weakly similar to No definition line found [C.elegans]
CGTTTTCTGA	594	20	6	-	43	1.84	Human protein-tyrosine phosphatase (HU-PP-1) mRNA, partial sequence
GGAAAAAAAA	595	97	8	-	187	1.84	Hepatocyte growth factor (hepapoietin A; scatter factor)
GGAAAAAAAA	596	97	8	-	187	1.84	ESTs, Highly similar to ATP SYNTHASE EPSILON CHAIN, MITOCHONDRIAL PRECURSOR [Bos taurus]
GAGGGAGTTT	597	548	162	-	1172	1.84	Ribosomal protein L27a
GACTCACTTT	598	156	27	-	315	1.84	Peptidylprolyl isomerase B (cyclophilin B)
GAGAACGGGG	599	33	7	-	67	1.85	ESTs, Highly similar to CORONIN [Dictyostelium discoideum]
TGGCTAGTGT	600	57	20	-	125	1.85	Human mRNA for proteasome subunit z, complete cds
CTGTCATTTG	601	20	5	-	42	1.85	PRE-MRNA SPLICING FACTOR SRP20
GTTCCCTGGC	602	320	98	-	690	1.85	Finkel-Biskis-Reilly murine sarcoma virus (FBR-MuSV) ubiquitously expressed (fox derived)
GCATTTAAAT	603	78	7	-	148	1.85	ELONGATION FACTOR 1-BETA

Table 4, cont.

ATCCACATCG	604	69	17	144	1.85	ESTs, Weakly similar to CASEIN KINASE I HOMOLOG HRR25 [Saccharomyces cerevisiae]
CTGCTGTGAT	605	29	6	59	1.85	Human mRNA for U1 small nuclear RNP-specific C protein
GTGACCTCCT	606	118	38	253	1.85	CYTOCHROME C OXIDASE POLYPEPTIDE VIII-LIVER/HEART PRECURSOR
GTGGACCCCA	607	47	9	97	1.86	Human siah binding protein 1 (SiahBP1) mRNA, partial cds
GACTAGTGCG	608	18	6	39	1.86	ESTs
TTATGGGATC	609	247	31	490	1.86	GUANINE NUCLEOTIDE-BINDING PROTEIN BETA SUBUNIT-LIKE PROTEIN 12.3
TTTCAGATTG	610	29	5	60	1.86	Human transcriptional coactivator PC4 mRNA, complete cds
GTCTGAGCTC	611	58	14	122	1.86	ESTs, Weakly similar to HYPOTHETICAL 15.4 KD PROTEIN C16C10.11 IN CHROMOSOME III [C.elegans]
CACACAATGT	612	22	9	49	1.86	Homo sapiens peroxisomal phytanoyl-CoA alpha-hydroxylase (PAHX) mRNA, complete cds
CACACAATGT	613	22	9	49	1.86	Cytochrome c oxidase subunit IV
ACCCCACCCA	614	26	6	55	1.86	H.sapiens mRNA for 1-acylglycerol-3-phosphate O-acyltransferase
GGAGGCAGGT	615	31	9	67	1.86	Homo sapiens chromosome 1p33-p34 beta-1,4-galactosyltransferase mRNA, complete cds
TCTCAATTCT	616	27	8	58	1.87	Cell division cycle 42 (GTP-binding protein, 25kD)
CTCTTCAGGA	617	19	6	40	1.87	Homo sapiens phosphomevalonate kinase mRNA, complete cds
CTGGGACTGC	618	18	7	40	1.87	Homo sapiens mRNA for follistain-related protein (FRP), complete cds
GCCCAGCAGG	619	26	8	57	1.87	ESTs
GCCCAGCAGG	620	26	8	57	1.87	ESTs
GGGCCAGGGG	621	44	18	98	1.87	ESTs
GGGGGACGGC	622	42	12	89	1.87	ESTs, Weakly similar to Y48E1B.1 [C.elegans]
ACTGGGTCTA	623	154	29	317	1.87	Non-metastatic cells 2, protein (NM23B) expressed in
GCCGAGGAAG	624	778	113	1570	1.87	Human mRNA for ribosomal protein S12
CAGATCTTTG	625	90	14	182	1.88	Ubiquitin A-52 residue ribosomal protein fusion product 1
AGGTTTCCTC	626	21	6	45	1.88	Homo sapiens mRNA for proteasome subunit p58, complete cds
CCGTCCAAGG	627	532	59	1058	1.88	Ribosomal protein S16
GTGGCGGGCG	628	81	21	174	1.88	Biliary glycoprotein
GTGGCGGGCG	629	81	21	174	1.88	Homo sapiens malignancy-associated protein mRNA, partial cds
GTGGCGGGCG	630	81	21	174	1.88	Homo sapiens mRNA for KIAA0565 protein, complete cds
GGCAAGAAGA	631	252	34	507	1.88	Ribosomal protein L27
TCTTTACTTG	632	23	8	49	1.88	Homo sapiens Arp2/3 protein complex subunit p21-Arc (ARC21) mRNA, complete cds
CTCCTCACCT	633	255	56	536	1.88	60S RIBOSOMAL PROTEIN L13A
CTCCTCACCT	634	255	56	536	1.88	Human Bak mRNA, complete cds
GCCTGTATGA	635	392	116	853	1.88	Ribosomal protein S24
GCTTTATTTG	636	560	147	1203	1.88	Human mRNA fragment encoding cytoplasmic actin, (isolated from cultured epidermal cells grown from human foreskin)
CTTAAGGATT	637	27	9	60	1.88	ESTs, Highly similar to transcription factor ARF6 chain B [M.musculus]
GGATTTGGCC	638	656	165	1401	1.88	Ribosomal protein, large P2
GGATTTGGCC	639	656	165	1401	1.88	Ribosomal protein S26
GGATTTGGCC	640	656	165	1401	1.88	Human mRNA for PIG-B, complete cds
TCCTCCCTCC	641	31	5	62	1.89	Human mRNA for proteasome subunit HsC7-I, complete cds
GGCCCTCTGA	642	46	9	98	1.89	Human peptidyl-prolyl isomerase and essential mitotic regulator (PIN1) mRNA, complete cds
TGGCTGTGTG	643	47	8	97	1.89	ESTs
AGACCAAAGT	644	38	6	79	1.89	DNAJ PROTEIN HOMOLOG 1

Table 4, cont.

ATGGCCAACT	645	28	12	-	64	1.89	ESTs
AGGAGCTGCT	646	81	12	-	165	1.89	ESTs
AGGAGCTGCT	647	81	12	-	165	1.89	Human mitochondrial NADH dehydrogenase-ubiquinone Fe-S protein 8, 23 kDa subunit precursor (NDUFS8) nuclear mRNA encoding mitochondrial protein, complete cds
TGTACCTGTA	648	245	8	-	473	1.90	Human alpha-tubulin mRNA, complete cds
GATCCCAACA	649	70	11	-	143	1.90	ATP synthase, H+ transporting, mitochondrial F1 complex, beta polypeptide
GGCCATCTCT	650	38	8	-	80	1.90	14-3-3 PROTEIN TAU
AGGTGCAGAG	651	26	9	-	58	1.90	Homo sapiens pescadillo mRNA, complete cds
GTGGCATCAC	652	32	7	-	68	1.90	ESTs, Weakly similar to C25A1.6 [C.elegans]
TGTGTTGAGA	653	1663	321	-	3487	1.90	Translation elongation factor 1-alpha-1
CTGAGACAAA	654	98	14	-	199	1.91	Basic transcription factor 3
GCAACGGGCC	655	54	6	-	108	1.91	Homo sapiens mRNA for brain acyl-CoA hydrolase, complete cds
GCTGGCTGGC	656	113	27	-	243	1.91	Homo sapiens chaperonin containing t-complex polypeptide 1, eta subunit (Ccth) mRNA, complete cds
GCCAAGATGC	657	55	11	-	116	1.91	ESTs
GCCAAGGGGC	658	28	8	-	61	1.91	Oxoglutarate dehydrogenase (lipoamide)
ACGGTGATGT	659	37	11	-	81	1.91	ESTs
CCCATCCGAA	660	353	77	-	753	1.91	Ribosomal protein L26
ACAAACTTAG	661	60	24	-	139	1.91	Human calmodulin mRNA, complete cds
GCCTCCTCCC	662	94	23	-	203	1.92	ESTs
GTGCCTGAGA	663	72	10	-	149	1.92	LAMIN A
TCCAATACTG	664	22	5	-	47	1.92	Human dynamitin mRNA, complete cds
GTGGTGCGTG	665	39	11	-	86	1.92	Homo sapiens X-ray repair cross-complementing protein 2 (XRCC2) mRNA, complete cds
AAGAAGCAGG	666	38	15	-	88	1.92	Homo sapiens unknown mRNA, complete cds
ACTTGGAGCC	667	42	13	-	95	1.92	Human calmodulin mRNA, complete cds
CCGTGGTCAC	668	88	15	-	185	1.92	H.sapiens mRNS for clathrin-associated protein
ACAGTGGGGA	669	65	21	-	146	1.92	Human (p23) mRNA, complete cds
ACAAACTGTG	670	69	22	-	154	1.92	H.sapiens mRNA for Sop2p-like protein
GTCTTAACTC	671	23	6	-	50	1.93	Homo sapiens Dim1p homolog (hdim1+) mRNA, complete cds
CTGTGCTCGG	672	34	11	-	77	1.93	ENOYL-COA HYDRATASE, MITOCHONDRIAL PRECURSOR
GTGGCCTGCA	673	22	5	-	46	1.93	ESTs, Weakly similar to K01G5.8 [C.elegans]
TGGTACACGT	674	100	43	-	236	1.93	Human calmodulin mRNA, complete cds
GTACTGTATG	675	23	9	-	54	1.93	ESTs
GTACTGTATG	676	23	9	-	54	1.93	Homo sapiens importin beta subunit mRNA, complete cds
GGCCAGGTGG	677	25	5	-	53	1.93	Homo sapiens calmodulin-stimulated phosphodiesterase PDE1B1 mRNA, complete cds
GGCCAGGTGG	678	25	5	-	53	1.93	Metallopeptidase 1 (33 kD)
AGGGAGAGGG	679	20	5	-	43	1.93	Homo sapiens forkhead protein FREAC-2 mRNA, complete cds
AGGGAGAGGG	680	20	5	-	43	1.93	Ferritin heavy chain
AGGGAGAGGG	681	20	5	-	43	1.93	UBIQUITIN CARBOXYL-TERMINAL HYDROLASE T
GTGGCAGGTG	682	100	19	-	213	1.93	Human mRNA for KIAA0340 gene, partial cds
TCTTGTGCAT	683	143	26	-	302	1.93	L-LACTATE DEHYDROGENASE M CHAIN
CCACACACCG	684	21	8	-	49	1.94	ESTs, Highly similar to HYPOTHETICAL 43.2 KD PROTEIN C34E10.1 IN CHROMOSOME III [Caenorhabditis elegans]
ACAAATCCTT	685	45	7	-	95	1.94	FK506-binding protein 1 (12kD)
GTGAGACCCC	686	45	11	-	98	1.94	No match

Table 4, cont.

AAAGCCAAGA	687	29	10	-	67	1.94	Electron-transfer-flavoprotein, beta polypeptide
CAAGGATCTA	688	27	12	-	65	1.94	Fibroblast growth factor receptor 2
TGAGGCCAGG	689	47	15	-	107	1.94	High mobility group box
TTTTGTGTGA	690	16	5	-	37	1.94	ESTs, Weakly similar to 50S RIBOSOMAL PROTEIN L20 [E.coli]
ACAGTCTTGC	691	17	6	-	38	1.94	CYTOCHROME P450 IVF3
ACAGTCTTGC	692	17	6	-	38	1.94	Human mRNA for KIAA0102 gene, complete cds
CCAGGCACGC	693	40	9	-	87	1.95	Human HXC-26 mRNA, complete cds
AGTTTCCCAA	694	40	21	-	100	1.95	Homo sapiens SULT1C sulfotransferase (SULT1C) mRNA, complete cds
CCAGTGGCCC	695	274	48	-	582	1.95	Ribosomal protein S9
GCCCCGCCCT	696	30	11	-	69	1.95	Homo sapiens chromosome 19, cosmid R32184
TCTCTACTAA	697	41	6	-	85	1.95	Tropomyosin 4 (fibroblast)
CGGCTTTTCT	698	32	9	-	71	1.95	Spectrin, beta, non-erythrocytic 1
TGGCCCCCGC	699	26	6	-	56	1.95	ESTs
TGGCCCCCGC	700	26	6	-	56	1.95	Human helix-loop-helix zipper protein mRNA
CTCCTGGGGC	701	48	6	-	101	1.95	ESTs
AAGGAGCTGG	702	16	5	-	37	1.96	ESTs, Highly similar to YME1 PROTEIN [Saccharomyces cerevisiae]
AAGGAGCTGG	703	16	5	-	37	1.96	ESTs
AAGGAGCTGG	704	16	5	-	37	1.96	Homo sapiens clone lambda MEN1 region unknown protein mRNA, complete cds
GGCTTTGATT	705	18	5	-	40	1.96	COATOMER BETA' SUBUNIT
ACTACCTTCA	706	27	8	-	61	1.96	ESTs, Weakly similar to B0334.4 [C.elegans]
CTGTGCATTT	707	33	11	-	75	1.96	Human 54 kDa protein mRNA, complete cds
ACTCCAAAAA	708	210	40	-	452	1.96	Human insulinoma rig-analog mRNA encoding DNA-binding protein, complete cds
ACTCCAAAAA	709	210	40	-	452	1.96	H.sapiens mRNA for transmembrane protein mp24
TCCTGCCCCA	710	72	14	-	155	1.96	Parathymosin
TCCTGCCCCA	711	72	14	-	155	1.96	Homo sapiens mRNA for KIAA0511 protein, partial cds
AAGCTGGAGG	712	56	15	-	125	1.96	Human translation initiation factor eIF3 p66 subunit mRNA, complete cds
GCACAAGAAG	713	90	19	-	185	1.96	ESTs
GAAACCGAGG	714	47	11	-	104	1.97	ESTs, Weakly similar to HYPOTHETICAL 16.8 KD PROTEIN IN SMY2-RPS101 INTERGENIC REGION [S.cerevisiae]
GAAACCGAGG	715	47	11	-	104	1.97	Human mRNA for KIAA0029 gene, partial cds
GCCCGCAAGC	716	16	5	-	36	1.97	H.sapiens HUNKI mRNA
CTTTCAGATG	717	44	12	-	98	1.97	Phosphofructokinase, platelet
GGGCGCTGTG	718	117	30	-	260	1.97	Homo sapiens mRNA for smallest subunit of ubiquinol-cytochrome c reductase, complete cds
GTATTCCCCT	719	36	8	-	79	1.97	Homo sapiens poly(A) binding protein II (PABP2) gene, complete cds
GTATTCCCCT	720	36	8	-	79	1.97	ESTs, Highly similar to elastin like protein [D.melanogaster]
CTGGCCATCG	721	19	6	-	43	1.98	ESTs
GTGGTGGACA	722	33	6	-	72	1.98	Human nicotinic acetylcholine receptor alpha6 subunit precursor, mRNA, complete cds
GTGGTGGACA	723	33	6	-	72	1.98	Homo sapiens mRNA for PBK1 protein
GTGGTGGACA	724	33	6	-	72	1.98	Breast cancer 1, early onset
CACCTAATTG	725	1247	410	-	2884	1.98	Tag matches mitochondrial sequence
GACCCCTGTC	726	18	6	-	41	1.98	Homo sapiens (clone s153) mRNA fragment
CCCTTAGCTT	727	47	21	-	114	1.98	Human mRNA for myosin regulatory light chain
CAGAGACGTG	728	30	9	-	68	1.98	Human dystroglycan (DAG1) mRNA, complete cds
ATGGCTGGTA	729	1064	174	-	2287	1.98	40S RIBOSOMAL PROTEIN S2
TCAGCCTTCT	730	46	14	-	108	1.99	Homo sapiens flotillin-1 mRNA, complete cds
TCGTAACGAG	731	23	9	-	54	1.99	ESTs

Table 4, cont.

GCGACGAGGC	732	178	17	-	371	1.99	60S RIBOSOMAL PROTEIN L38
GCGGGGTACC	733	59	17	-	133	1.99	Human mRNA for pM5 protein
TCCTTCTCCA	734	58	12	-	128	1.99	ALPHA-ACTININ 1, CYTOSKELETAL ISOFORM
CAGTCTCTCA	735	107	16	-	229	1.99	Ribosomal protein S10
ACCCTTCCCT	736	56	12	-	124	1.99	ESTs, Weakly similar to VON EBNER'S GLAND PROTEIN PRECURSOR [H.sapiens]
ACCCTTCCCT	737	56	12	-	124	1.99	Signal sequence receptor, beta
TGAGTGGTCA	738	20	7	-	47	1.99	ESTs, Highly similar to HYPOTHETICAL 13.6 KD PROTEIN IN NUP170-ILS1 INTERGENIC REGION [Saccharomyces cerevisiae]
GACAATGCCA	739	48	11	-	107	1.99	Human mRNA for ATP synthase gamma-subunit (L-type), complete cds
ATCTTTCTGG	740	80	15	-	176	2.00	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide
AGCTGTCCCC	741	23	5	-	50	2.00	Tag matches mitochondrial sequence
TCTTCCAGGA	742	52	11	-	114	2.00	Human ribosomal protein L10 mRNA, complete cds
GTGCCTAGGA	743	29	9	-	67	2.00	ESTs
TGGACCCCCC	744	26	6	-	57	2.00	ESTs, Weakly similar to K04G2.2 [C.elegans]
ACCTGTATCC	745	158	24	-	341	2.00	INTERFERON-INDUCIBLE PROTEIN 1-8U
ACCTGCTGGT	746	17	6	-	40	2.00	Homo sapiens clone 23675 mRNA sequence
AGTCTGATGT	747	39	5	-	84	2.00	ESTs, Weakly similar to weak similarity to rat TEGT protein [C.elegans]
TCTCTACCCA	748	71	27	-	169	2.00	Amyloid beta (A4) precursor-like protein 2
TGATTAAGGT	749	26	6	-	58	2.00	HEAT SHOCK FACTOR PROTEIN 1
CAGCAGAAGC	750	191	75	-	459	2.01	Homo sapiens 4F5rel mRNA, complete cds
TCCCTATTAA	751	5970	987	-	12977	2.01	No match
GTGGAGGTGC	752	42	6	-	91	2.01	Human 100 kDa coactivator mRNA, complete cds
AAGATCCCCG	753	63	15	-	142	2.01	Homo sapiens DNA sequence from cosmid ICK0721Q on chromosome 6.
GAGCGGCCTC	754	29	9	-	68	2.01	Human ORF mRNA, complete cds
AACTACATAG	755	21	9	-	50	2.02	ESTs
GTAAGATTTG	756	33	9	-	76	2.02	Human 150 kDa oxygen-regulated protein ORP150 mRNA, complete cds
AGCCTGCAGA	757	65	17	-	147	2.02	Homo sapiens chromosome 19, cosmid R33729
GGACCACTGA	758	498	174	-	1162	2.02	Ribosomal protein L3
TTCAATAAAA	759	377	51	-	813	2.02	TRANSCOBALAMIN I PRECURSOR
TTCAATAAAA	760	377	51	-	813	2.02	Ribosomal protein, large, P1
CGATGGTCCC	761	55	9	-	120	2.02	Human B-cell receptor associated protein (hBAP) mRNA, partial cds
CATTTGTAAT	762	142	23	-	309	2.02	Tag matches mitochondrial sequence
CCTGAGCCCG	763	60	14	-	135	2.03	ESTs, Weakly similar to ALBUMIN B-32 PROTEIN [Zea mays]
TGAGGCCTCT	764	29	6	-	65	2.03	ESTs
AAGAGTTACG	765	17	8	-	43	2.03	ESTs, Highly similar to 50S RIBOSOMAL PROTEIN L2 [Bacillus stearothermophilus]
GAATCCAACT	766	46	6	-	100	2.03	ESTs
AGGGGCGCAG	767	29	8	-	67	2.03	Human SH3-containing protein EEN mRNA, complete cds
GCTTAGAAGT	768	31	6	-	69	2.03	HEAT SHOCK PROTEIN HSP 90-ALPHA
AAGTCATTCA	769	31	10	-	74	2.03	Homo sapiens NADH-ubiquinone oxidoreductase subunit CI-B14 mRNA, complete cds
AAGTCATTCA	770	31	10	-	74	2.03	H.sapiens mRNA for prcc protein
TACCCCACCC	771	57	17	-	132	2.03	ESTs
TACCCCACCC	772	57	17	-	132	2.03	Human zinc finger protein (MAZ) mRNA
CCTAGCTGGA	773	511	132	-	1172	2.03	PEPTIDYL-PROLYL CIS-TRANS ISOMERASE A
TCGTCTTTAT	774	126	18	-	275	2.04	40S RIBOSOMAL PROTEIN S7
GGTTTGGCTT	775	70	14	-	156	2.04	UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX 11 KD PROTEIN PRECURSOR

Table 4, cont.

TAGGATGGGG	776	88	28	-	207	2.04	Sodium/potassium-transporting ATPase beta-3 subunit
GTGCATCCCG	777	43	16	-	105	2.04	Casein kinase 2, beta polypeptide
CAGCGCTGCA	778	37	11	-	87	2.04	Human CDC37 homolog mRNA, complete cds
GGGAGCCCCT	779	55	12	-	125	2.04	ESTs, Highly similar to BETA-ARRESTIN 2 [Homo sapiens]
GGGAGCCCCT	780	55	12	-	125	2.04	ESTs
GAAGATGTGG	781	58	6	-	125	2.04	Homo sapiens clone 23967 unknown mRNA, partial cds
CCTACCACAG	782	21	9	-	52	2.05	ESTs, Highly similar to GOLIATH PROTEIN [Drosophila melanogaster]
TGCTAAAAAA	783	26	9	-	61	2.06	Myosin, heavy polypeptide 9, non-muscle
CACAGAGTCC	784	28	7	-	64	2.06	Low density lipoprotein-related protein-associated protein 1 (alpha-2-macroglobulin receptor-associated protein 1
GGGCCAATAA	785	30	8	-	70	2.06	Untitled
GCCTGCTGGG	786	220	49	-	503	2.07	Phospholipid hydroperoxide glutathione peroxidase
ACTGCTTGCC	787	52	12	-	118	2.07	S-ADENOSYLMETHIONINE SYNTHETASE GAMMA FORM
ACTGCTTGCC	788	52	12	-	118	2.07	H.sapiens mRNA for Sop2p-like protein
CGGTTACTGT	789	81	20	-	187	2.07	Homo sapiens NADH:ubiquinone oxidoreductase NDUFS6 subunit mRNA, nuclear gene encoding mitochondrial protein, complete cds
AACCCGGGAG	790	179	50	-	420	2.07	Homo sapiens KIAA0408 mRNA, complete cds
AACCCGGGAG	791	179	50	-	420	2.07	Cytokine receptor family II, member 4
AACCCGGGAG	792	179	50	-	420	2.07	H.sapiens mRNA for delta 4-3-oxosteroid 5 beta-reductase
ATTAACAAAG	793	98	18	-	220	2.07	Guanine nucleotide binding protein (G protein), alpha stimulating activity polypeptide 1
TTCAGTGCCC	794	18	6	-	43	2.07	ESTs, Weakly similar to GLUCOSE-6-PHOSPHATASE [Rattus norvegicus]
CCGTGCTCAT	795	51	18	-	123	2.07	ESTs, Highly similar to ADIPOCYTE P27 PROTEIN [Mus musculus]
ATCCCTCAGT	796	78	24	-	184	2.07	Activating transcription factor 4 (tax-responsive enhancer element B67)
TACCATCAAT	797	664	194	-	1985	2.07	Glyceraldehyde-3-phosphate dehydrogenase
TGCACCACAG	798	34	14	-	84	2.08	Homo sapiens signal peptidase complex 18 kDa subunit mRNA, partial cds
GAACCCTGGG	799	46	9	-	104	2.08	ESTs
GCCGTGTCCG	800	542	60	-	1185	2.08	Human ribosomal protein S6 mRNA, complete cds
ATAGAGGCAA	801	26	7	-	65	2.08	Human mRNA for KIAA0026 gene, complete cds
ATTGTTTATG	802	83	11	-	184	2.08	Human non-histone chromosomal protein HMG-17 mRNA, complete cds
TAATAAAGGT	803	229	46	-	523	2.09	40S RIBOSOMAL PROTEIN S8
GGGATCAAGG	804	26	7	-	61	2.09	ESTs, Weakly similar to coded for by C. elegans cDNA yk157f8.5 [C.elegans]
CAAGGGCTTG	805	28	8	-	88	2.09	ESTs, Highly similar to RAS-RELATED PROTEIN RAP-1B [Homo sapiens; Bos taurus]
TGGTGTTGAG	806	828	147	-	1876	2.09	Human DNA sequence from clone 1033B10 on chromosome 6p21.2-21.31.
GAGTGAGTGA	807	19	8	-	48	2.09	ESTs, Weakly similar to C44C1.2 gene product [C.elegans]
GTGGCGCACA	808	42	9	-	98	2.09	Human mRNA for KIAA0072 gene, partial cds
ATGATCCGGA	809	22	5	-	52	2.10	ATPase, Ca++ transporting, cardiac muscle, slow twitch 2
AACCTGGGAG	810	108	37	-	263	2.10	Human DNA fragmentation factor-45 mRNA, complete cds
AACCTGGGAG	811	108	37	-	263	2.10	Homo sapiens mRNA for KIAA0563 protein, complete cds
TGCTTCATCT	812	53	9	-	120	2.10	Homo sapiens androgen receptor associated protein 24 (ARA24) mRNA, complete cds
ATAATTCTTT	813	205	37	-	467	2.10	Ribosomal protein S29
GTTCAGCTGT	814	41	9	-	95	2.10	Voltage-dependent anion channel 2
GGGAAGTCAC	815	22	5	-	50	2.10	Human FX protein mRNA, complete cds
GGGTGCTTGG	816	26	8	-	63	2.10	Human mRNA for ORF, Xq terminal portion
CAGTTACTTA	817	52	11	-	120	2.10	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, beta polypeptide
GCGAAACCCC	818	207	70	-	506	2.10	Human G protein-coupled receptor (STRL22) mRNA, complete cds

Table 4, cont.

GCCTTCCAAT	819	85	11	-	191	2.11	P68 PROTEIN
CCCCCTGGAT	820	485	33	-	1056	2.11	Cell division cycle 2-like 1 (PITSLRE proteins)
GACCTCCTGC	821	21	5	-	49	2.12	Homo sapiens mRNA for kinesin-like DNA binding protein, complete cds
GACCTCCTGC	822	21	5	-	49	2.12	Human SH3 domain-containing proline-rich kinase (sprk) mRNA, complete cds
CAGCAGTAGC	823	23	6	-	55	2.12	H.sapiens mRNA for 218kD Mi-2 protein
TTCATTATAA	824	47	8	-	108	2.12	Prothymosin alpha
CCCCCACCTA	825	64	15	-	150	2.12	INTESTINAL MEMBRANE A4 PROTEIN
GGTGGATGTG	826	30	6	-	69	2.12	Homo sapiens methyl-CpG binding protein MBD3 (MBD3) mRNA, complete cds
TCTGGTTTGT	827	41	5	-	91	2.12	Homo sapiens mRNA for integral membrane protein Tmp21-I (p23)
TCTGGTTTGT	828	41	5	-	91	2.12	THYMOSIN BETA-10
CGCCTGTAAT	829	48	8	-	111	2.13	CDC21 HOMOLOG
TCCTGCTGCC	830	45	6	-	101	2.13	ESTs
TCCTGCTGCC	831	45	6	-	101	2.13	ESTs, Weakly similar to F46F6.1 [C.elegans]
GTGTGGTGGT	832	27	6	-	64	2.13	Homo sapiens mRNA for GDP dissociation inhibitor beta
TGATGTCCAC	833	10	5	-	27	2.14	ESTs
CCAGGAGGAA	834	222	77	-	551	2.14	HEAT SHOCK COGNATE 71 KD PROTEIN
GTGAAGCCCC	835	42	9	-	99	2.14	No match
GGGAGCCCGG	836	32	7	-	75	2.15	Homo sapiens herpesvirus entry protein B (HVEB) mRNA, complete cds
GCCATCCCCT	837	64	14	-	150	2.15	Tag matches mitochondrial sequence
CAGTTGGTTG	838	28	8	-	69	2.15	Homo sapiens mRNA for E1B-55kDa-associated protein
ATCCATCTGT	839	21	9	-	54	2.15	H.sapiens hnRNP-E2 mRNA
GCCAGGAAGC	840	32	6	-	75	2.15	ESTs, Weakly similar to C01A2.5 [C.elegans]
TCCAGCCCCT	841	32	9	-	78	2.15	ESTs, Weakly similar to T08G11.1 [C.elegans]
GCCCCCCACT	842	24	6	-	58	2.15	Human MAP kinase activated protein kinase 2 mRNA, complete cds
TGTCTGTGGT	843	18	5	-	45	2.15	H.sapiens BAT1 mRNA for nuclear RNA helicase (DEAD family)
TCCCGTACAT	844	258	37	-	592	2.15	No match
GTGGTGGGCA	845	61	12	-	144	2.15	Cholinergic receptor, nicotinic, delta polypeptide
GTGGTGGGCA	846	61	12	-	144	2.15	Isovaleryl Coenzyme A dehydrogenase
GTGGTGGGCA	847	61	12	-	144	2.15	Homo sapiens josephin MJD1 mRNA, complete cds
CTGTTAGTGT	848	54	13	-	130	2.16	MALATE DEHYDROGENASE, CYTOPLASMIC
CTCTCACCCT	849	68	28	-	175	2.16	Ribonuclease/angiogenin inhibitor
TGCTGGTGTG	850	30	8	-	74	2.16	Human mRNA, clone HH109 (screened by the monoclonal antibody of insulin receptor substrate-1 (IRS-1))
CTAAGACTTC	851	1455	317	-	3462	2.16	Tag matches mitochondrial sequence
GGAAGGACAG	852	39	5	-	90	2.16	ATPase, H+ transporting, lysosomal (vacuolar proton pump) 31kD
GAAGTGTGTC	853	23	9	-	60	2.16	ESTs, Highly similar to HYPOTHETICAL 37.2 KD PROTEIN C12C2.09C IN CHROMOSOME I [Schizosaccharomyces pombe]
GTACCCGGAC	854	33	9	-	81	2.17	ESTs, Weakly similar to W08E3.1 [C.elegans]
CCTCCCTGAT	855	35	10	-	86	2.17	Homo sapiens dynamin (DNM) mRNA, complete cds
TCATCTTCAA	856	19	5	-	46	2.17	CALRETICULIN PRECURSOR
TCATCTTCAA	857	19	5	-	46	2.17	ESTs
TCATCTTCAA	858	19	5	-	46	2.17	RAB6, member RAS oncogene family
ATGTACTCTG	859	38	6	-	89	2.17	IMP (inosine monophosphate) dehydrogenase 2
CGCCGGAACA	860	648	123	-	1530	2.17	Ribosomal protein L4
AAGGGAGGGT	861	78	14	-	184	2.17	Human phosphotyrosine independent ligand p62 for the Lck SH2 domain mRNA, complete cds
GAAAAAAAAA	862	112	12	-	255	2.17	Cell division cycle 10 (homologous to CDC10 of S. cerevisiae

Table 4, cont.

AAACTCTGTG	863	27	6	64	2.18	Homo sapiens p120 catenin isoform 1A (CTNND1) mRNA, alternatively spliced, complete cds
ACACACGCAA	864	22	8	56	2.18	ESTs
CCGCCGAAGT	865	50	7	116	2.18	Ribosomal protein L12
TGTGCTAAAT	866	169	46	415	2.18	60S RIBOSOMAL PROTEIN L34
CGACCGTGGC	867	24	6	57	2.18	ESTs
GCCTGGGCTG	868	44	18	114	2.18	ESTs
GCCTGGGCTG	869	44	18	114	2.18	Homo sapiens molybdopterin synthase sulfurylase (MOCS3) mRNA, complete cds
AAAGTCAGAA	870	24	12	65	2.19	Ubiquinol-cytochrome c reductase core protein II
TGGAGCGCTA	871	31	5	71	2.19	ESTs, Weakly similar to PUTATIVE MITOCHONDRIAL CARRIER C16C10.1 [C.elegans]
GAAATGATGA	872	70	14	167	2.19	Homo sapiens mRNA for c-myc binding protein, complete cds
TGTCGCTGGG	873	73	14	173	2.19	C4/C2 activating component of Ra-reactive factor
GCCCCTGCCT	874	39	6	91	2.19	Homo sapiens DNA-binding protein (CROC-1B) mRNA, complete cds
GCCCCTGCCT	875	39	6	91	2.19	Glutathione S-transferase M4
CAGGCCTGGC	876	20	7	50	2.19	ESTs
CAGGCCTGGC	877	20	7	50	2.19	ESTs
GCAAAAAAAA	878	153	35	371	2.20	No match
AGCCACCACG	879	33	8	81	2.20	Human mRNA for KIAA0149 gene, complete cds
GAGGAAGAAG	880	52	16	130	2.20	Homologue of mouse tumor rejection antigen gp96
CAGCTGTAGT	881	20	9	54	2.20	Human mRNA for KIAA0174 gene, complete cds
TCTTCTCCCT	882	40	10	99	2.20	Human mRNA for hepatoma-derived growth factor, complete cds
TACATTCTGT	883	30	7	74	2.20	Myeloid cell leukemia sequence 1 (BCL2-related)
GGGAAACCCC	884	39	11	98	2.21	ESTs, Weakly similar to HYPOTHETICAL 68.7 KD PROTEIN ZK757.1 IN CHROMOSOME III [C.elegans]
AGCCACTGCA	885	67	8	155	2.21	Homo sapiens mRNA for 26S proteasome subunit p55, complete cds
TAGTTGAAGT	886	55	13	136	2.21	UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX 14 KD PROTEIN
GCCAAGTTTG	887	17	5	43	2.21	Human mRNA for proteasome subunit p112, complete cds
GGCGGCTGCA	888	38	9	89	2.21	Excision repair cross-complementing rodent repair deficiency, complementation group 1 (includes overlapping antisense sequence)
AAAAAAAAAA	889	469	38	1076	2.21	H.sapiens mRNA for sodium-phophate transport system 1
AAAAAAAAAA	890	469	38	1076	2.21	Homo sapiens GPI-linked anchor protein (GFRA1) mRNA, complete cds
AAAAAAAAAA	891	469	38	1076	2.21	Enolase 1, (alpha)
AAAAAAAAAA	892	469	38	1076	2.21	Calcium channel, voltage-dependent, P/Q type, alpha 1A subunit
TGTTCCACTC	893	18	5	46	2.21	Homo sapiens CD39L2 (CD39L2) mRNA, complete cds
CTCGGTGATG	894	30	10	76	2.22	H.sapiens mRNA for ras-related GTP-binding protein
CTTCTCAGGG	895	17	5	43	2.22	ESTs, Highly similar to PUTATIVE CYSTEINYL-TRNA SYNTHETASE C29E6.06C [Schizosaccharomyces pombe]
GGTAGCCCAC	896	16	5	40	2.22	ESTs
GGGTTTTTAT	897	65	7	150	2.22	Homo sapiens dbpB-like protein mRNA, complete cds
CCTGTAACCC	898	39	12	99	2.23	Human translation initiation factor eIF-2alpha mRNA, 3'UTR
GAAACAAGAT	899	58	5	133	2.23	Phosphoglycerate kinase 1
GATGAGTCTC	900	71	18	175	2.23	Homo sapiens proteasome subunit XAPC7 mRNA, complete cds
GGCCCTAGGC	901	43	6	101	2.23	H.sapiens ERF-2 mRNA
TGGCCCCACC	902	440	59	1041	2.23	Pyruvate kinase, muscle
CAGCGCGCCC	903	66	5	152	2.23	ESTs
AGGCGAGATC	904	91	27	231	2.24	Homo sapiens proteasome subunit XAPC7 mRNA, complete cds

Table 4, cont.

GCGGGGTGGA	905	64	12	-	155	2.24	H.sapiens ERF-1 mRNA 3' end
GGGGCCCCCT	906	21	6	-	54	2.24	Homo sapiens mRNA for NA14 protein
AAGGAACTTG	907	24	8	-	61	2.24	ESTs
AAGGAACTTG	908	24	8	-	61	2.24	Homo sapiens clone 24655 mRNA sequence
AATTGCAAGC	909	18	5	-	47	2.24	COFILIN, NON-MUSCLE ISOFORM
CCTGTGATCC	910	66	22	-	171	2.25	No match
CCCCGCCAAG	911	66	11	-	159	2.25	Human adult heart mRNA for neutral calponin, complete cds
CTCAACAGCA	912	60	12	-	147	2.25	Human translation initiation factor 3 47 kDa subunit mRNA, complete cds
AAGGTAGCAG	913	56	17	-	143	2.25	ADENYLYL CYCLASE-ASSOCIATED PROTEIN 1
AAGCCAGCCC	914	78	5	-	180	2.25	Protein kinase C substrate 80K-H
CAGCCTTGGA	915	21	5	-	52	2.25	ESTs, Weakly similar to siah binding protein 1 [H.sapiens]
TTTGCTCTCC	916	24	8	-	61	2.25	Vinculin
CAACATTCCT	917	41	14	-	106	2.26	Dopachrome tautomerase (dopachrome delta-isomerase, tyrosine-related protein 2)
TACTAGTCCT	918	77	13	-	187	2.26	HEAT SHOCK PROTEIN HSP 90-ALPHA
GACTCTGGTG	919	59	6	-	139	2.26	Homo sapiens chromosome 19, cosmid R29381
GACTCTGGTG	920	59	6	-	139	2.26	40S RIBOSOMAL PROTEIN S15A
GTGGCTCACG	921	102	16	-	248	2.26	Homo sapiens KIAA0414 mRNA, partial cds
GTGGCTCACG	922	102	16	-	248	2.26	Human Tax1 binding protein mRNA, partial cds
GTGGCGGGCA	923	71	16	-	177	2.27	H.sapiens mRNA for urea transporter
GTGGCGGGCA	924	71	16	-	177	2.27	Homo sapiens mRNA for KIAA0472 protein, partial cds
CCTGTGGTCC	925	86	18	-	215	2.27	No match
TACAGCACGG	926	27	6	-	68	2.27	Homo sapiens microsomal glutathione S-transferase 3 (MGST3) mRNA, complete cds
GTGGCACCTG	927	20	5	-	51	2.27	ESTs, Highly similar to NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG PRECURSOR [Xenopus laevis]
TACACGTGAG	928	40	14	-	103	2.27	ESTs, Weakly similar to GOLIATH PROTEIN [Drosophila melanogaster]
TCAGGCATTT	929	69	24	-	180	2.27	ESTs, Highly similar to RAS-RELATED PROTEIN RAB-1A [H.sapiens]
TTCACAAAGG	930	25	7	-	63	2.27	PROTEASOME ZETA CHAIN
TTCTTGTGGC	931	245	54	-	610	2.27	Ribosomal protein S11
TCCCTATTAG	932	91	14	-	220	2.27	No match
TACAAGAGGA	933	208	49	-	521	2.27	Ribosomal protein L6
TCAGACGCAG	934	344	78	-	862	2.28	Prothymosin alpha
CAGGATCCAG	935	35	6	-	86	2.28	Human putative tumor suppressor (SNC6) mRNA, complete cds
TCTGTACACC	936	55	11	-	135	2.28	Ribosomal protein S11
GAAGCAGGAC	937	352	54	-	856	2.28	COFILIN, NON-MUSCLE ISOFORM
GCGCCGCCCC	938	27	5	-	68	2.28	ESTs, Moderately similar to nuclear autoantigen [H.sapiens]
CCCTCCTGGG	939	69	23	-	181	2.29	ESTs
TGGGCGCCTT	940	35	6	-	85	2.29	Uroporphyrinogen decarboxylase
GTGGTACAGG	941	121	35	-	312	2.29	Homo sapiens microtubule-based motor (HsKIFC3) mRNA, complete cds
GTGGTACAGG	942	121	35	-	312	2.29	ESTs
GGTGAGACCT	943	93	43	-	255	2.29	Prostatic binding protein
GAGATCCGCA	944	59	16	-	153	2.30	INTERFERON GAMMA UP-REGULATED I-5111 PROTEIN PRECURSOR
TTGGCAGCCC	945	48	5	-	115	2.30	Ribosomal protein L27a
GCCTTTCCCT	946	22	8	-	59	2.30	APOPTOSIS REGULATOR BCL-X
GGAGTGGACA	947	190	29	-	465	2.30	60S RIBOSOMAL PROTEIN L18
TTATGGGGAG	948	29	6	-	74	2.30	H factor (complement)-like 1
TTATGGGGAG	949	29	6	-	74	2.30	TRANSFORMATION-SENSITIVE PROTEIN IEF SSP 3521

Table 4, cont.

GAGTGGGGGC	950	43	9	- 108	2.30	ESTs, Highly similar to LYSOSOMAL PRO-X CARBOXYPEPTIDASE PRECURSOR [Homo sapiens]
GTGGCACGTG	951	192	36	- 479	2.30	No match
CTGGGCGTGT	952	126	41	- 331	2.31	ESTs
TTGGGGTTTC	953	1243	255	- 3123	2.31	Ferritin heavy chain
GGCTGGGCCT	954	93	14	- 229	2.31	Clathrin, light polypeptide (Lcb)
GGCTGGGCCT	955	93	14	- 229	2.31	EST
CCTGTTCTCC	956	28	8	- 73	2.31	ESTs
GTGTCTCATC	957	26	6	- 67	2.31	ESTs
GTGTCTCATC	958	26	6	- 67	2.31	Enolase 1, (alpha)
ACGATTGATG	959	23	6	- 60	2.31	ESTs, Highly similar to HYPOTHETICAL 27.5 KD PROTEIN IN SPX19-GCR2 INTERGENIC REGION [Saccharomyces cerevisiae]
TTGTTGTTGA	960	75	20	- 194	2.31	Calmodulin 1 (phosphorylase kinase, delta)
TGGCCTCCCC	961	49	9	- 122	2.32	H.sapiens mRNA for rho GDP-dissociation inhibitor 1
ATCGGGCCCG	962	51	19	- 136	2.32	ESTs, Weakly similar to zinc finger protein [H.sapiens]
GCCGCCATCA	963	45	8	- 111	2.33	Human protein disulfide isomerase-related protein P5 mRNA, partial cds
GTGCTGGACC	964	63	15	- 162	2.33	Human mRNA for proteasome activator hPA28 subunit beta, complete cds
TTGTAATCGT	965	208	59	- 540	2.33	Human mRNA for ornithine decarboxylase antizyme, ORF 1 and ORF 2
TAATGGTAAC	966	30	5	- 75	2.33	Homo sapiens nuclear-encoded mitochondrial cytochrome c oxidase Va subunit mRNA, complete cds
AACGACCTCG	967	156	6	- 369	2.33	Homo sapiens clone 24703 beta-tubulin mRNA, complete cds
GCCTGCACCC	968	18	7	- 49	2.34	Human neuronal olfactomedin-related ER localized protein mRNA, partial cds
GCCTGCACCC	969	18	7	- 49	2.34	ESTs
AAGGTGGAGG	970	609	156	- 2051	2.34	60S RIBOSOMAL PROTEIN L18A
AAGGAGATGG	971	467	132	- 1226	2.34	Ribosomal protein L31
CAGTTCTCTG	972	41	9	- 105	2.34	Human BTK region clone ftp-3 mRNA
GTGAAACCTC	973	111	38	- 297	2.35	Homo sapiens intrinsic factor-B12 receptor precursor, mRNA, complete cds
TAGGTTGTCT	974	546	104	- 1386	2.35	TRANSLATIONALLY CONTROLLED TUMOR PROTEIN
CCTGTGACAG	975	61	8	- 150	2.35	Human mRNA for KIAA0106 gene, complete cds
CTCATAAGGA	976	572	118	- 1463	2.35	Tag matches mitochondrial sequence
GGTGGCTTTG	977	23	8	- 61	2.35	Homo sapiens NADH:ubiquinone oxidoreductase B12 subunit mRNA, nuclear gene encoding mitochondrial protein, complete cds
GCTCAGCTGG	978	171	28	- 432	2.36	Eukaryotic translation elongation factor 1 delta (guanine nucleotide exchange protein)
GGCCCTGAGC	979	141	14	- 348	2.36	Human RNA polymerase II subunit (hsRPB10) mRNA, complete cds
TCTGCTAAAG	980	53	5	- 130	2.36	High-mobility group (nonhistone chromosomal) protein 1
TCTGCTAAAG	981	53	5	- 130	2.36	ESTs
AGCCCCACAA	982	18	5	- 46	2.37	ESTs
CTGAGTCTCC	983	80	9	- 198	2.37	Guanine nucleotide binding protein (G protein), alpha inhibiting activity polypeptide 2
TGCTTTGGGA	984	53	14	- 139	2.37	ESTs, Weakly similar to No definition line found [C.elegans]
CCTGTCCTGC	985	60	7	- 149	2.37	ESTs, Moderately similar to GTP-binding protein-associated protein [M.musculus]
GGGGAAATCG	986	708	98	- 1772	2.37	THYMOSIN BETA-10
TCTGCCTGGG	987	48	15	- 130	2.37	ESTs, Weakly similar to orf, len: 159, CAI: 0.12 [S.cerevisiae]
CAATAAACTG	988	97	12	- 242	2.37	PROTEIN TRANSLATION FACTOR SUI1 HOMOLOG
GAGTCTGAGG	989	24	9	- 68	2.37	U1 snRNP 70K protein
GTGGCAGGCG	990	87	16	- 223	2.37	Human pancreatic zymogen granule membrane protein GP-2 mRNA, complete cds
GTGGCAGGCG	991	87	16	- 223	2.37	Nuclear factor of kappa light polypeptide gene enhancer in B-cells 2 (p49/p100)
CGAGGGGCCA	992	186	33	- 480	2.38	Human non-muscle alpha-actinin mRNA, complete cds

Table 4, cont.

GTGGGGGGAG	993	19	5	-	49	2.38	Human DNA sequence from cosmid F0811 on chromosome 6. Contains Daxx, BING1, Tapasin, RGL2, KE2, BING4, BING5, ESTs and CpG islands
GAGTGGCTAT	994	28	8	-	75	2.38	Homo sapiens KIAA0419 mRNA, complete cds
GAGTGGCTAT	995	28	8	-	75	2.38	Homo sapiens mRNA for GDP dissociation inhibitor beta
GTAGACTCAC	996	17	5	-	46	2.38	LARGE PROLINE-RICH PROTEIN BAT2
AGGGAAAGAG	997	27	7	-	72	2.39	Human G10 homolog (edg-2) mRNA, complete cds
AGGGAAAGAG	998	27	7	-	72	2.39	Homo sapiens mRNA for KIAA0632 protein, partial cds
CCCATCGTCC	999	3108	714	-	8145	2.39	Tag matches mitochondrial sequence
TCGCCGCGAC	1000	34	8	-	90	2.40	No match
TGTCCTGGTT	1001	150	39	-	398	2.40	CYCLIN-DEPENDENT KINASE INHIBITOR 1
CTTTTTGTGC	1002	42	6	-	107	2.40	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, beta polypeptide
ATAAATTGGG	1003	23	8	-	82	2.40	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit b, isoform 1
TATCACTCTG	1004	21	6	-	57	2.40	Human male-enhanced antigen mRNA (Mea), complete cds
GTGGTGGGCG	1005	61	9	-	156	2.40	No match
CCACTACACT	1006	38	6	-	98	2.41	Human TNF-related apoptosis inducing ligand TRAIL mRNA, complete cds
TGACCCCACA	1007	29	11	-	81	2.41	ESTs, Weakly similar to F25H5.h [C.elegans]
TGATTTCACT	1008	803	132	-	2064	2.41	EST
TGATTTCACT	1009	803	132	-	2064	2.41	Tag matches mitochondrial sequence
GGCTCCCACT	1010	142	36	-	379	2.41	HEAT SHOCK PROTEIN HSP 90-BETA
CCTGTGTGTG	1011	32	8	-	82	2.41	ESTs
AATCCTGTGG	1012	514	135	-	1377	2.42	Ribosomal protein L8
AGGAGCAAAG	1013	43	9	-	112	2.42	Human mRNA for NADPH-flavin reductase, complete cds
CCTTTGAACA	1014	43	7	-	111	2.42	Human Chromosome 16 BAC clone CIT987SK-A-61E3
GTGGGGCTAG	1015	30	8	-	81	2.42	H.sapiens mRNA for protein phosphatase 5
AGGGTGAAAC	1016	29	5	-	75	2.43	Human splicing factor SRp30c mRNA, complete cds
CCTCAGGATA	1017	270	72	-	728	2.43	ESTs
CCTCAGGATA	1018	270	72	-	728	2.43	Tag matches mitochondrial sequence
TTCCACTAAC	1019	55	12	-	147	2.44	Human plectin (PLEC1) mRNA, complete cds
CCCCCGTGAA	1020	86	18	-	228	2.44	Homo sapiens interleukin-1 receptor-associated kinase (IRAK) mRNA, complete cds
TGTGCTCGGG	1021	107	35	-	295	2.44	Human mRNA for KIAA0088 gene, partial cds
AAGCCTTGCT	1022	20	6	-	54	2.44	ESTs
TGTTCATCAT	1023	40	15	-	114	2.45	ESTs, Weakly similar to neuroendocrine-specific protein C [H.sapiens]
AACTAACAAA	1024	86	24	-	234	2.45	Ubiquitin A-52 residue ribosomal protein fusion product 1
GCTGTTGCGC	1025	158	33	-	419	2.45	40S RIBOSOMAL PROTEIN S20
GGATGTGAAA	1026	45	7	-	118	2.45	Antigen identified by monoclonal antibodies 12E7, F21 and O13
ACTGGTACGT	1027	34	8	-	90	2.45	Homo sapiens F1Fo-ATPase synthase f subunit mRNA, complete cds
TTGTATTCCA	1028	16	5	-	45	2.45	H.sapiens mRNA for alpha 4 protein
GGCTGGGGGC	1029	437	48	-	1124	2.46	Human profilin mRNA, complete cds
CCACTGCACT	1030	925	181	-	2460	2.47	Thyroid autoantigen 70kD (Ku antigen)
CCACTGCACT	1031	925	181	-	2460	2.47	Enhancer of zeste (Drosophila) homolog 1
CCACTGCACT	1032	925	181	-	2460	2.47	CD19 antigen
CCACTGCACT	1033	925	181	-	2460	2.47	Human clone 23732 mRNA, partial cds
CCACTGCACT	1034	925	181	-	2460	2.47	Annexin II (lipocortin II)
CCACTGCACT	1035	925	181	-	2460	2.47	Alkaline phosphatase, placental (Regan isozyme)
CCACTGCACT	1036	925	181	-	2460	2.47	Homo sapiens clone 24760 mRNA sequence
CCACTGCACT	1037	925	181	-	2460	2.47	Homo sapiens carbonic anhydrase precursor (CA 12) mRNA, complete cds

Table 4, cont.

CCACTGCACT	1038	925	181	-	2460	2.47	Homo sapiens methyl-CpG binding protein MBD4 (MBD4) mRNA, complete cds
CCACTGCACT	1039	925	181	-	2460	2.47	Phosphodiesterase 4C, cAMP-specific (dunce (Drosophila)-homolog phosphodiesterase E1)
CCACTGCACT	1040	925	181	-	2460	2.47	Human SNRPN mRNA, 3' UTR, partial sequence
CCACTGCACT	1041	925	181	-	2460	2.47	Homo sapiens brachyury variant A (TBX1) mRNA, complete cds
CCACTGCACT	1042	925	181	-	2460	2.47	H.sapiens beta glucuronidase pseudogene
CCACTGCACT	1043	925	181	-	2460	2.47	G PROTEIN-ACTIVATED INWARD RECTIFIER POTASSIUM CHANNEL 4
CACTTGCCCT	1044	109	21	-	290	2.47	ESTs, Highly similar to ACETYL-COENZYME A SYNTHETASE (Escherichia coli)
CACTTGCCCT	1045	109	21	-	290	2.47	ESTs, Highly similar to NADH-UBIQUINONE OXIDOREDUCTASE B22 SUBUNIT (Bos taurus)
GCAAGCCAAC	1046	100	17	-	284	2.47	Tag matches mitochondrial sequence
TAGATAATGG	1047	49	5	-	126	2.47	Homo sapiens clone 24703 beta-tubulin mRNA, complete cds
TCGAAGCCCC	1048	251	60	-	682	2.47	Tag matches mitochondrial sequence
AGAAAAAAAA	1049	115	9	-	294	2.48	Enolase 1, (alpha)
AGAAAAAAAA	1050	115	9	-	294	2.48	Human mRNA for KIAA0099 gene, complete cds
GGCGCCTCCT	1051	66	9	-	172	2.48	Eukaryotic translation initiation factor 4A (eIF-4A) isoform 1
GGCGCCTCCT	1052	66	9	-	172	2.48	TRANSALDOLASE
TAAACTGTTT	1053	29	7	-	79	2.48	ESTs
TAAACTGTTT	1054	29	7	-	79	2.48	40S RIBOSOMAL PROTEIN S14
GGCCTTTTTT	1055	36	6	-	95	2.48	Human mRNA for histone H1x, complete cds
GGCCTTTTTT	1056	36	6	-	95	2.48	Homo sapiens mRNA for KIAA0529 protein, partial cds
GCGACAGCTC	1057	44	5	-	115	2.48	60S RIBOSOMAL PROTEIN L24
CCCACACTAC	1058	57	17	-	159	2.49	Human signal-transducing guanine nucleotide-binding regulatory (G) protein beta subunit mRNA, complete cds
AGCAGATCAG	1059	390	65	-	1034	2.49	S100 calcium-binding protein A10 (annexin II ligand, calpactin I, light polypeptide (p11))
GCATAGGCTG	1060	90	15	-	240	2.49	ELONGATION FACTOR TU, MITOCHONDRIAL PRECURSOR
GAGGCCGACC	1061	25	9	-	72	2.49	Basigin
AAATGCCACA	1062	42	6	-	110	2.49	ESTs, Weakly similar to neuroendocrine-specific protein C [H.sapiens]
AGCCCTACAA	1063	754	208	-	2089	2.49	Tag matches mitochondrial sequence
TTGGTGAAGG	1064	399	57	-	1053	2.50	Human thymosin beta-4 mRNA, complete cds
CCGGGCCCAG	1065	46	9	-	125	2.50	Homo sapiens mRNA for TRIP6 (thyroid receptor interacting protein)
TTCATACACC	1066	772	125	-	2055	2.50	Tag matches mitochondrial sequence
GCAGCCATCC	1067	790	96	-	2072	2.50	Ribosomal protein L28
GCCGGGTGGG	1068	668	126	-	1798	2.50	Basigin
GCTCCCAGAC	1069	53	9	-	142	2.50	Homo sapiens mRNA for synaptogyrin 2
AGCCACCGTG	1070	39	6	-	105	2.51	No match
TCAGCTGGCC	1071	16	6	-	47	2.51	Human nuclear factor NF90 mRNA, complete cds
GGGGGCGCCT	1072	22	6	-	62	2.52	Adenine nucleotide translocator 3 (liver)
CGGCCCAACG	1073	59	14	-	161	2.52	H.sapiens mRNA for arginine methyltransferase, splice variant, 1262 bp
TGGCCATCTG	1074	65	14	-	177	2.52	ESTs, Weakly similar to N-methyl-D-aspartate receptor glutamate-binding chain [R.norvegicus]
CCTCCCCCGT	1075	59	11	-	159	2.52	Homo sapiens breakpoint cluster region protein 1 (BCRG1) mRNA, complete cds
ACTTGTTCGC	1076	27	6	-	73	2.52	ESTs
AAGACTGGCT	1077	30	6	-	81	2.52	ESTs, Highly similar to Surf-4 protein [M.musculus]
AGCACATTTG	1078	42	5	-	112	2.53	ESTs, Highly similar to deduced protein product shows significant homology to coactosin from Dictyostelium discoideum [H.sapiens]
GTGAAGGCAG	1079	467	83	-	1265	2.53	Ribosomal protein S3A

Table 4, cont.

CAATAAATGT	1080	227	43	-	620	2.54	Ribosomal protein L37
GCCAGGGCGG	1081	46	5	-	121	2.54	ESTs, Highly similar to HYPOTHETICAL 52.8 KD PROTEIN T05E11.5 IN CHROMOSOME IV [Caenorhabditis elegans]
GTGTAATAAG	1082	57	9	-	154	2.54	Heterogeneous nuclear ribonucleoprotein A2/B1
TTCTGCACTG	1083	25	6	-	70	2.54	Collagen, type I, alpha-2
TTCTGCACTG	1084	25	6	-	70	2.54	ESTs
GTGAAACCCC	1085	1352	514	-	3963	2.55	Myelin oligodendrocyte glycoprotein (alternative products)
GTGAAACCCC	1086	1352	514	-	3963	2.55	Dihydrolipoamide branched chain transacylase (E2 component of branched chain keto acid dehydrogenase complex)
GTGAAACCCC	1087	1352	514	-	3963	2.55	Human mRNA for platelet-activating factor acetylhydrolase 2, complete cds
GTGAAACCCC	1088	1352	514	-	3963	2.55	GRANULOCYTE-MACROPHAGE COLONY-STIMULATING FACTOR RECEPTOR ALPHA CHAIN PRECURSOR
GTGAAACCCC	1089	1352	514	-	3963	2.55	Thymopoietin
GTGAAACCCC	1090	1352	514	-	3963	2.55	Basic fibroblast growth factor (bFGF) receptor (shorter form)
GTGAAACCCC	1091	1352	514	-	3963	2.55	Homo sapiens mRNA for KIAA0794 protein, partial cds
GTGAAACCCC	1092	1352	514	-	3963	2.55	Homo sapiens RNA polymerase I subunit hRPA39 mRNA, complete cds
GTGAAACCCC	1093	1352	514	-	3963	2.55	Homo sapiens mRNA for KIAA0701 protein, partial cds
GTGAAACCCC	1094	1352	514	-	3963	2.55	Homo sapiens mRNA for MAX.3 cell surface antigen
GTGAAACCCC	1095	1352	514	-	3963	2.55	Homo sapiens mRNA for KIAA0706 protein, complete cds
GTGAAACCCC	1096	1352	514	-	3963	2.55	Homo sapiens deoxyribonuclease II mRNA, complete cds
GTGAAACCCC	1097	1352	514	-	3963	2.55	Homo sapiens clone 24758 mRNA sequence
GTGAAACCCC	1098	1352	514	-	3963	2.55	Kangai 1 (suppression of tumorigenicity 6, prostate; CD82 antigen (R2 leukocyte antigen, antigen detected by monoclonal and antibody IA4))
GTGAAACCCC	1099	1352	514	-	3963	2.55	Leptin (murine obesity homolog)
GACACCTCCT	1100	45	7	-	122	2.55	ESTs, Weakly similar to TIP49 [R.norvegicus]
GACGTGTGGG	1101	94	6	-	247	2.56	H2AZ histone
GCAAAACCCC	1102	162	46	-	461	2.56	Homo sapiens tumor necrosis factor superfamily member LIGHT mRNA, complete cds
TACCAGTGTA	1103	46	6	-	124	2.56	Heat shock 60 kD protein 1 (chaperonin)
CCCCTCCCCA	1104	30	11	-	90	2.58	Chromosome 22q13 BAC Clone CIT987SK-384D8 complete sequence
GGTGATGAGG	1105	35	8	-	98	2.58	Homo sapiens BC-2 protein mRNA, complete cds
GTGTGTAAAA	1106	27	6	-	76	2.59	H.sapiens CDM mRNA
GGCTCCTCGA	1107	41	11	-	117	2.59	Homo sapiens tapasin (NGS-17) mRNA, complete cds
AAAAGAAACT	1108	62	12	-	174	2.60	POLYADENYLATE-BINDING PROTEIN
CAGCGCACAG	1109	22	5	-	64	2.60	ESTs
CTGGGAGAGG	1110	35	11	-	102	2.60	ESTs
GAAAAATGGT	1111	340	58	-	943	2.60	Laminin receptor (2H5 epitope)
ATCACGCCCT	1112	192	26	-	527	2.61	Tag matches mitochondrial sequence
TAGCTCTATG	1113	107	43	-	323	2.61	ATPase, Na+/K+ transporting, alpha 1 polypeptide
GTATTGGCCT	1114	21	7	-	61	2.61	Human p76 mRNA, complete cds
CCCGACGTGC	1115	58	20	-	171	2.62	ESTs, Highly similar to NADH-UBIQUINONE OXIDOREDUCTASE B9 SUBUNIT [Bos taurus]
GAAGTTATGA	1116	32	7	-	89	2.62	T-COMPLEX PROTEIN 1, ALPHA SUBUNIT
TAAAAAAAAA	1117	108	7	-	290	2.63	ESTs
TAAAAAAAAA	1118	108	7	-	290	2.63	Ubiquitin-conjugating enzyme E2A (RAD6 homolog)
TAAAAAAAAA	1119	108	7	-	290	2.63	Homo sapiens protein kinase (BUB1) mRNA, complete cds
GCCGCCCTGC	1120	71	13	-	199	2.63	Acyl-Coenzyme A dehydrogenase, very long chain
TTTGGGGCTG	1121	78	30	-	234	2.63	Human mRNA for proton-ATPase-like protein, complete cds

Table 4, cont.

GTGGCAGGCA	1122	86	16	-	245	2.63	No match
GGCTGTACCC	1123	79	16	-	225	2.63	CYSTEINE-RICH PROTEIN
AGCAGGGCTC	1124	128	17	-	353	2.63	ESTs, Highly similar to PNG gene [H.sapiens]
AAGAAGATAG	1125	152	10	-	412	2.64	60S RIBOSOMAL PROTEIN L23A
TCTGGGGACG	1126	27	7	-	78	2.64	Human translational initiation factor 2 beta subunit (eIF-2-beta) mRNA, complete cds
GCTAGGTTTA	1127	80	9	-	220	2.65	Tag matches mitochondrial sequence
TGGTGACAGT	1128	32	6	-	91	2.65	Homo sapiens histone H2A.F/Z variant (H2AV) mRNA, complete cds
TTACCATATC	1129	196	46	-	566	2.65	Human mRNA for ribosomal protein L39, complete cds
GTGGCGGGTG	1130	59	9	-	165	2.65	No match
TGGATCCTAG	1131	28	7	-	81	2.66	Homo sapiens NADH:ubiquinone oxidoreductase NDUFS3 subunit mRNA, nuclear gene encoding mitochondrial protein, complete cds
GGGTTTGAAC	1132	22	7	-	64	2.66	Homo sapiens SKB1Hs mRNA, complete cds
AATGCAGGCA	1133	83	9	-	231	2.67	S-adenosylhomocysteine hydrolase
ACATCGTAGG	1134	30	10	-	90	2.67	ESTs
AACGCTGCCT	1135	59	10	-	167	2.67	Human APRT gene for adenine phosphoribosyltransferase
TGGAGGTGGG	1136	20	6	-	58	2.68	ESTs
TGCCTGCTCC	1137	21	8	-	64	2.68	ESTs
CTTCCAGCTA	1138	358	87	-	1050	2.69	Annexin II (lipocortin II)
GTAAGTGTAC	1139	80	8	-	223	2.69	ESTs
GTAAGTGTAC	1140	80	8	-	223	2.69	Tag matches mitochondrial sequence
GTGTCTCGCA	1141	40	6	-	112	2.70	Annexin XI (56kD autoantigen)
ATCCGGCGCC	1142	114	14	-	321	2.70	Homo sapiens RNA polymerase II transcription factor SIII p18 subunit mRNA, complete cds
TGCCTGCACC	1143	232	61	-	688	2.70	Cystatin C (amyloid angiopathy and cerebral hemorrhage)
TTCCTATTAA	1144	42	7	-	121	2.72	ESTs
CAGGAGTTCA	1145	91	23	-	270	2.72	Homo sapiens Arp2/3 protein complex subunit p34-Arc (ARC34) mRNA, complete cds
GTCTGCGTGC	1146	51	5	-	143	2.72	Proteasome component C2
GAAATACAGT	1147	264	50	-	769	2.72	ESTs
GAAATACAGT	1148	264	50	-	769	2.72	Cathepsin D (lysosomal aspartyl protease)
TGAGCCCGGC	1149	36	8	-	106	2.74	ESTs, Highly similar to LATENT TRANSFORMING GROWTH FACTOR BETA BINDING PROTEIN 1 PRECURSOR [Rattus norvegicus]
GTGGTGTGTG	1150	46	8	-	134	2.74	Homo sapiens NF-AT4c mRNA, complete cds
GTGGTGTGTG	1151	46	8	-	134	2.74	Acid phosphatase, prostate
TCACCCACAC	1152	383	111	-	1167	2.76	Ribosomal protein L17
TCACCCACAC	1153	383	111	-	1167	2.76	ESTs, Weakly similar to !!!! ALU SUBFAMILY J WARNING ENTRY !!!! [H.sapiens]
CTGGATCTGG	1154	65	12	-	190	2.76	Glycogen phosphorylase B (brain form)
GAAGATGTGT	1155	95	24	-	287	2.77	ESTs, Highly similar to HYPOTHETICAL 6.3 KD PROTEIN ZK652.2 IN CHROMOSOME III [Caenorhabditis elegans]
CGGATAACCA	1156	53	6	-	153	2.78	Human cell cycle protein p38-2G4 homolog (hG4-1) mRNA, complete cds
TCAGAAGGTG	1157	38	5	-	111	2.78	ESTs, Weakly similar to RNA-binding protein [H.sapiens]
GAGAAACCCC	1158	95	22	-	288	2.78	Human mRNA for KIAA0134 gene, complete cds
GAGAAACCCC	1159	95	22	-	288	2.78	H.sapiens F11 mRNA
GAGAAACCCC	1160	95	22	-	288	2.78	Human mRNA for KIAA0159 gene, complete cds
CTCGTTAAGA	1161	32	6	-	95	2.80	Human calmodulin mRNA, complete cds
TTGGAGATCT	1162	93	20	-	279	2.80	Human NADH:ubiquinone oxidoreductase MLRQ subunit mRNA, complete cds
GAGGTCCCTG	1163	65	12	-	193	2.81	PROTEASOME IOTA CHAIN
TTCCGCGTGC	1164	50	5	-	146	2.81	Homo sapiens lysyl hydroxylase isoform 3 (PLOD3) mRNA, complete cds

Table 4, cont.

CAGCCCAACC	1165	64	8	-	187	2.81	Homo sapiens eukaryotic translation initiation factor 3 subunit (p42) mRNA, complete cds
GTGGCTCACA	1166	104	9	-	303	2.81	Adenosine A2b receptor
TAGAAAGGCA	1167	31	6	-	92	2.82	H.sapiens ERF-2 mRNA
TAAGTAGCAA	1168	33	7	-	102	2.83	ESTs, Weakly similar to putative [M.musculus]
GGTGAGACAC	1169	128	25	-	389	2.83	Adenine nucleotide translocator 3 (liver)
CCCATCGTCT	1170	39	5	-	116	2.83	No match
CCGATCACCG	1171	59	14	-	182	2.83	Human translational initiation factor 2 beta subunit (eIF-2-beta) mRNA, complete cds
GAATCGGTTA	1172	43	10	-	133	2.83	Homo sapiens NADH-ubiquinone oxidoreductase 15kDa subunit mRNA, complete cds
AACCCAGGAG	1173	110	11	-	323	2.84	No match
TTTTGAAGCA	1174	33	15	-	108	2.85	Homo sapiens hepatitis B virus X interacting protein (XIP) mRNA, complete cds
CACAGGCAAA	1175	40	8	-	122	2.85	Human mRNA for KIAA0005 gene, complete cds
TCAGCTTCAC	1176	30	7	-	93	2.85	Human mRNA for KIAA0359 gene, complete cds
TCAGCTTCAC	1177	30	7	-	93	2.85	Human putative G-protein (GP-1) mRNA, complete cds
GAGGGCCGGT	1178	61	10	-	185	2.85	ESTs, Highly similar to HISTONE H2A [Cairina moschata]
CCCCAGCCAG	1179	320	74	-	988	2.86	Ribosomal protein S3
GTGGTGGGTG	1180	59	5	-	176	2.86	Human RACH1 (RACH1) mRNA, complete cds
CTGCCAAGTT	1181	100	27	-	314	2.87	Homo sapiens mRNA for zyxin
GAGAAACCCT	1182	46	12	-	144	2.87	Homo sapiens mRNA, chromosome 1 specific transcript KIAA0506
GAGAAACCCT	1183	46	12	-	144	2.87	Vitamin D (1,25- dihydroxyvitamin D3) receptor
ACTAACACCC	1184	544	132	-	1694	2.87	Tag matches mitochondrial sequence
TTTTGGGGGC	1185	37	7	-	112	2.88	ESTs
TTTTGGGGGC	1186	37	7	-	112	2.88	Human mRNA for proton-ATPase-like protein, complete cds
GTGAAACCCA	1187	43	15	-	140	2.88	No match
GCTTTCATTG	1188	27	12	-	89	2.89	Homo sapiens clone 23967 unknown mRNA, partial cds
GTGGCACGCA	1189	33	6	-	101	2.89	No match
GGGTCAAAAG	1190	52	14	-	165	2.89	HISTONE H3.3
GGGGGTCACC	1191	61	9	-	186	2.90	ATP SYNTHASE LIPID-BINDING PROTEIN P1 PRECURSOR
GTGAAACCCT	1192	664	198	-	2130	2.91	Carboxypeptidase M
GTGAAACCCT	1193	664	198	-	2130	2.91	H.sapiens mRNA for laminin
GTGAAACCCT	1194	664	198	-	2130	2.91	GC-RICH SEQUENCE DNA-BINDING FACTOR
GTGAAACCCT	1195	664	198	-	2130	2.91	Homo sapiens mRNA for KIAA0596 protein, partial cds
GTGAAACCCT	1196	664	198	-	2130	2.91	Homo sapiens clone 23605 mRNA sequence
GTGAAACCCT	1197	664	198	-	2130	2.91	Formyl peptide receptor 1
AGTTGAAATT	1198	20	6	-	84	2.91	ESTs
AGAATCGCTT	1199	74	11	-	226	2.92	Homo sapiens coatomer protein (COPA) mRNA, complete cds
AGGTCAAGAG	1200	20	7	-	65	2.92	No match
CTAACCAGAC	1201	43	11	-	136	2.93	ANGIOTENSIN-CONVERTING ENZYME PRECURSOR, SOMATIC
GGGATGGCAG	1202	38	5	-	115	2.93	VALYL-TRNA SYNTHETASE
AGACCCACAA	1203	162	39	-	512	2.93	Tag matches mitochondrial sequence
TCGAAGAACC	1204	50	7	-	155	2.94	CD63 antigen (melanoma 1 antigen)
TGAAATAAAA	1205	71	8	-	214	2.95	Nucleophosmin (nucleolar phosphoprotein B23, numatrin)
ACTGAGGTGC	1206	34	9	-	109	2.95	Homo sapiens FGF-1 intracellular binding protein (FIBP) mRNA, complete cds
ACTCAGAAGA	1207	50	12	-	160	2.95	ESTs, Highly similar to NADH-UBIQUINONE OXIDOREDUCTASE AGGG SUBUNIT PRECURSOR [Bos taurus]
GAACACATCC	1208	440	113	-	1414	2.96	Ribosomal protein L19
AACTAATACT	1209	67	6	-	203	2.96	ESTs, Weakly similar to !!!! ALU SUBFAMILY J WARNING ENTRY !!!! [H.sapiens]

Table 4, cont.

AGATGTGTGG	1210	30	8	-	98	2.96	Hydroxyacyl-Coenzyme A dehydrogenase/3-ketoacyl-Coenzyme A thiolase/enoyl-Coenzyme A hydratase (trifunctional protein), beta subunit
GTGGTGTGCA	1211	27	8	-	89	2.97	Homo sapiens RNA transcript from U17 small nucleolar RNA host gene, variant U17HG-AB
GGCGTCCTGG	1212	55	9	-	172	2.98	ESTs, Weakly similar to No definition line found [C.elegans]
CCTGCAATCC	1213	47	11	-	152	2.98	No match
GCCTGGCCAT	1214	57	14	-	184	2.99	GUANINE NUCLEOTIDE-BINDING PROTEIN BETA SUBUNIT-LIKE PROTEIN 12.3
GCCTGGCCAT	1215	57	14	-	184	2.99	ESTs, Moderately similar to SULFATED SURFACE GLYCOPROTEIN 185 [Volvox carteri]
GCTGCCCTTG	1216	134	14	-	415	2.99	Human alpha-tubulin mRNA, 3' end
GCTGCCCTTG	1217	134	14	-	415	2.99	Human alpha-tubulin mRNA, complete cds
GCCAGCCCAG	1218	90	12	-	281	3.00	Human transcriptional corepressor hKAP1/TIF1B mRNA, complete cds
TCCTATTAAG	1219	160	34	-	515	3.00	ESTs
ATTGTGCCAC	1220	34	8	-	110	3.00	No match
CCATTGCACT	1221	237	58	-	773	3.02	Ataxia telangiectasia mutated (includes complementation groups A, C and D)
GCACCTCAGC	1222	38	8	-	122	3.02	ESTs
TTGGTCAGGC	1223	129	24	-	419	3.05	Calcium modulating ligand
TTGGTCAGGC	1224	129	24	-	419	3.05	Human melanoma antigen recognized by T-cells (MART-1) mRNA
GGGCCCCGCA	1225	30	6	-	98	3.05	Human mRNA for KIAA0123 gene, partial cds
GTGGCACACA	1226	70	15	-	228	3.06	Homo sapiens AIBC1 (AIBC1) mRNA, complete cds
GTGGCACACA	1227	70	15	-	228	3.06	Homo sapiens mRNA for MEGF8, partial cds
TTGGCCAGGC	1228	346	87	-	1149	3.07	Human cytochrome P450-IIB (hIIB3) mRNA, complete cds
TTGGCCAGGC	1229	346	87	-	1149	3.07	Homo sapiens X-ray repair cross-complementing protein 2 (XRCC2) mRNA, complete cds
TTGGCCAGGC	1230	346	87	-	1149	3.07	Homo sapiens oligodendrocyte-specific protein (OSP) mRNA, complete cds
TTGGCCAGGC	1231	346	87	-	1149	3.07	MHC class II transactivator
TTGGCCAGGC	1232	346	87	-	1149	3.07	Fc fragment of IgA, receptor for
TTGGCCAGGC	1233	346	87	-	1149	3.07	Protein kinase, interferon-inducible double stranded RNA dependent
TTGGCCAGGC	1234	346	87	-	1149	3.07	Zinc finger protein 157 (HZF22)
GTCACTGCCT	1235	20	5	-	68	3.08	Homo sapiens mRNA for Ribosomal protein kinase B (RSK-B)
GCCACCCCGT	1236	61	8	-	197	3.09	Glucose-6-phosphate dehydrogenase
TCCCTATAAG	1237	107	17	-	347	3.09	No match
CCTGTAATCC	1238	1302	453	-	4484	3.10	Breast cancer 2, early onset
CCTGTAATCC	1239	1302	453	-	4484	3.10	Integrin, beta 3 (platelet glycoprotein IIIa, antigen CD61)
CCTGTAATCC	1240	1302	453	-	4484	3.10	Transcription factor 1, hepatic; LF-B1, hepatic nuclear factor (HNF1), albumin proximal factor
CCTGTAATCC	1241	1302	453	-	4484	3.10	Homo sapiens interferon induced tetratricopeptide protein IFI60 (IFIT4) mRNA, complete cds
CCTGTAATCC	1242	1302	453	-	4484	3.10	H.sapiens RBQ-3 mRNA
CCTGTAATCC	1243	1302	453	-	4484	3.10	Human hVps41p (HVPS41) mRNA, complete cds
CCTGTAATCC	1244	1302	453	-	4484	3.10	Human TNF-alpha converting enzyme precursor, mRNA, alternatively spliced, complete cds
CCTGTAATCC	1245	1302	453	-	4484	3.10	Homo sapiens mRNA for KIAA0526 protein, complete cds
CCTGTAATCC	1246	1302	453	-	4484	3.10	Homo sapiens melastatin 1 (MLSN1) mRNA, complete cds
CCTGTAATCC	1247	1302	453	-	4484	3.10	Homo sapiens clone 23716 mRNA sequence
CCTGTAATCC	1248	1302	453	-	4484	3.10	Homo sapiens mRNA for KIAA0538 protein, partial cds

Table 4, cont.

CCTGTAATCC	1249	1302	453	-	4484	3.10	HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, E E*0101/E*0102 ALPHA CHAIN PRECURSOR
CCTGTAATCC	1250	1302	453	-	4484	3.10	Homo sapiens decoy receptor 2 mRNA, complete cds
CCTGTAATCC	1251	1302	453	-	4484	3.10	CATHEPSIN S PRECURSOR
CCTGTAATCC	1252	1302	453	-	4484	3.10	Homo sapiens type 6 nucleoside diphosphate kinase NM23-H6 (NM23-H6) mRNA, complete cds
CCTGTAATCC	1253	1302	453	-	4484	3.10	5' nucleotidase (CD73)
CCTGTAATCC	1254	1302	453	-	4484	3.10	Homo sapiens mRNA, chromosome 1 specific transcript KIAA0508
CCTGTAATCC	1255	1302	453	-	4484	3.10	H.sapiens mRNA for p85 beta subunit of phosphatidyl-inositol-3-kinase
CCTGTAATCC	1256	1302	453	-	4484	3.10	Interleukin 12 receptor, beta-2
TCCCCGTACA	1257	3918	290	-	12438	3.10	No match
GTCACACCAC	1258	30	9	-	104	3.11	ESTs
GTCACACCAC	1259	30	9	-	104	3.11	Prothymosin alpha
ATGGCAAGGG	1260	56	9	-	182	3.11	ESTs, Weakly similar to !!!! ALU SUBFAMILY J WARNING ENTRY !!!! [H.sapiens]
CTGTTGGCAT	1261	111	27	-	372	3.11	Ribosomal protein L21
CTAGCCTCAC	1262	623	161	-	2105	3.12	Actin, gamma 1
AGTGCAAGAC	1263	57	10	-	187	3.12	Tag matches mitochondrial sequence
CCTGTAGTCC	1264	231	67	-	791	3.13	No match
TTTTCTGAAA	1265	66	12	-	218	3.13	Thioredoxin
CTCCCCTGCC	1266	62	9	-	203	3.14	Capping protein (actin filament), gelsolin-like
TCTCTTTTTC	1267	32	6	-	108	3.14	H.sapiens tissue specific mRNA
GCGGACGAGG	1268	35	8	-	118	3.14	Homo sapiens TFAR19 mRNA, complete cds
GCGGACGAGG	1269	35	8	-	118	3.14	Human tip associating protein (TAP) mRNA, complete cds
GGAGTCATTG	1270	56	12	-	190	3.16	Human mRNA for proteasome subunit HsC10-II, complete cds
GTAGCAGGTG	1271	67	21	-	233	3.17	Homo sapiens cargo selection protein TIP47 (TIP47) mRNA, complete cds
CGCAAGCTGG	1272	65	13	-	221	3.17	LAMIN A
GTGAAACCCG	1273	36	11	-	126	3.18	No match
AGGTCAGGAG	1274	359	133	-	1274	3.18	Major histocompatibility complex, class II, DR beta 5
AGGTCAGGAG	1275	359	133	-	1274	3.18	Human mRNA for KIAA0331 gene, complete cds
AGGTCAGGAG	1276	359	133	-	1274	3.18	Human mRNA for KIAA0226 gene, complete cds
GAATGCAGTT	1277	13	5	-	45	3.18	ESTs
GAATGCAGTT	1278	13	5	-	45	3.18	ESTs
GAATGCAGTT	1279	13	5	-	45	3.18	ESTs
GTGAGCCCAT	1280	77	21	-	269	3.21	HEAT SHOCK PROTEIN HSP 90-BETA
GTAATCCTGC	1281	109	23	-	375	3.22	Tag matches ribosomal RNA sequence
TGAAGTAACA	1282	31	7	-	108	3.22	PROTEIN TRANSLATION FACTOR SUI1 HOMOLOG
TGCCTGTAAT	1283	59	15	-	206	3.22	ISLET AMYLOID POLYPEPTIDE PRECURSOR
GTAGCATAAA	1284	28	6	-	95	3.23	Human ubiquitin gene, complete cds
CCGTGGTCGT	1285	67	9	-	224	3.23	Fibrillarin
ATGAAACCCC	1286	67	24	-	240	3.23	Homo sapiens mRNA expressed in osteoblast, complete cds
AAGATTGGTG	1287	81	13	-	275	3.25	CD9 antigen
ATCCGTGCCC	1288	35	11	-	124	3.25	Human calmodulin mRNA, complete cds
CCCTTCACTG	1289	16	5	-	58	3.26	ESTs, Moderately similar to !!!! ALU SUBFAMILY J WARNING ENTRY !!!! [H.sapiens]
CCCTTCACTG	1290	16	5	-	58	3.26	ESTs
CAGCTGGGGC	1291	54	6	-	183	3.26	Polypyrimidine tract binding protein (hnRNP I) (alternative products)
CAGGCCCCAC	1292	109	17	-	370	3.26	Human mRNA for calgizzarin, complete cds
TGTTTATCCT	1293	25	7	-	89	3.26	-

Table 4, cont.

TAACCAATCA	1294	52	14	-	184	3.26	Human Rab5c-like protein mRNA, complete cds
CACCTGTAGT	1295	32	5	-	110	3.27	Ribosomal protein L5
TACCCTAAAA	1296	103	16	-	351	3.27	Human kpni repeat mrna (cdna clone pcd-kpni-4), 3' end
TACCCTAAAA	1297	103	16	-	351	3.27	Homo sapiens mRNA for KIAA0675 protein, complete cds
TACCCTAAAA	1298	103	16	-	351	3.27	Human Line-1 repeat mRNA with 2 open reading frames
TGCCTCTGCG	1299	175	83	-	655	3.28	Human platelet-endothelial tetraspan antigen 3 mRNA, complete cds
GCAAAACCCT	1300	81	19	-	284	3.28	No match
AAGGACCTTT	1301	115	18	-	396	3.28	ESTs
CTGGCGCCGA	1302	39	9	-	138	3.30	ESTs, Weakly similar to F35G12.9 [C.elegans]
GAAGCTTTGC	1303	133	15	-	454	3.30	HEAT SHOCK PROTEIN HSP 90-ALPHA
GCTCCGAGCG	1304	57	6	-	195	3.30	Ribosomal protein S16
TTGCCCAGGC	1305	69	21	-	251	3.30	Cell division cycle 42 (GTP-binding protein, 25kD)
TTGCCCAGGC	1306	69	21	-	251	3.30	Human brain mRNA homologous to 3'UTR of human CD24 gene, partial sequence
ACCCACGTCA	1307	55	9	-	189	3.31	Jun B proto-oncogene
GCTCCACTGG	1308	29	8	-	103	3.31	Mannose-6-phosphate receptor (cation dependent)
TTTAACGGCC	1309	142	18	-	489	3.31	Tag matches mitochondrial sequence
CTTGTAATCC	1310	71	11	-	248	3.32	ESTs, Moderately similar to !!!! ALU SUBFAMILY J WARNING ENTRY !!!! [H.sapiens]
CACTTTTGGG	1311	47	8	-	165	3.33	ESTs
CCGGGTGATG	1312	92	20	-	325	3.33	Human copper transport protein HAH1 (HAH1) mRNA, complete cds
GGGGTAAGAA	1313	62	6	-	213	3.33	Prostatic binding protein
TGACTGGCAG	1314	49	7	-	172	3.34	CD59 antigen p18-20 (antigen identified by monoclonal antibodies 16.3A5, EJ16, EJ30, EL32 and G344)
CAATGTGTTA	1315	47	17	-	176	3.39	H.sapiens mRNA for NADH dehydrogenase
GGCTCGGGAT	1316	74	6	-	257	3.40	CALPAIN 1, LARGE
TGCCTGTAGT	1317	71	15	-	258	3.40	Hum ORF (CEI5) mRNA, 3' flank
CGCCGCCGGC	1318	807	148	-	2906	3.42	Human ribosomal protein L35 mRNA, complete cds
GGTGGGGAGA	1319	68	6	-	239	3.44	Human chromosome 17q21 mRNA clone LF113
GTAAAACCCT	1320	24	8	-	90	3.44	No match
GGCTCCTGGC	1321	100	9	-	354	3.44	Homo sapiens b(2)gcn homolog mRNA, complete cds
AGTAGGTGGC	1322	53	5	-	188	3.46	Tag matches mitochondrial sequence
GGAGGTGGGG	1323	126	19	-	456	3.48	Granulin
CCTTTGGCTA	1324	27	5	-	100	3.49	ESTs, Highly similar to 40S RIBOSOMAL PROTEIN S27 [Rattus norvegicus]
AGAAAGATGT	1325	74	11	-	268	3.50	Annexin I (lipocortin I)
AGAACAAAAC	1326	75	8	-	271	3.52	Proliferation-associated gene A (natural killer-enhancing factor A)
AACTAAAAAA	1327	110	9	-	396	3.53	Ubiquitin A-52 residue ribosomal protein fusion product 1
ATTGCACCAC	1328	38	5	-	138	3.53	Human transglutaminase mRNA, 3' untranslated region
GATCCCAACT	1329	389	27	-	1402	3.54	H.sapiens mRNA for metallothionein isoform 2
GATCCCAACT	1330	389	27	-	1402	3.54	Human mRNA for metallothionein from cadmium-treated cells
CACTACTCAC	1331	356	99	-	1361	3.54	Tag matches mitochondrial sequence
CTGTACAGAC	1332	132	20	-	487	3.55	Homo sapiens beta 2 gene
TACCCTAGAA	1333	43	5	-	159	3.58	Estrogen receptor
GTAAAACCCC	1334	57	8	-	213	3.58	Tumor necrosis factor receptor 2 (75kD)
GTAAAACCCC	1335	57	8	-	213	3.58	Homo sapiens mRNA for KIAA0632 protein, partial cds
GTAAAACCCC	1336	57	8	-	213	3.58	Homo sapiens protease-activated receptor 4 mRNA, complete cds
CTGAGAGCTG	1337	32	9	-	125	3.61	Homo sapiens growth-arrest-specific protein (gas) mRNA, complete cds
GGCTGGTCTG	1338	57	8	-	211	3.62	ESTs
ACGCAGGGAG	1339	360	29	-	1334	3.63	HEAT SHOCK PROTEIN HSP 90-ALPHA

Table 4, cont.

GCCCTCGGCC	1340	44	5	-	165	3.63	Homo sapiens mRNA for protein phosphatase 2C gamma
CTCCCTTGCC	1341	20	5	-	78	3.64	ESTs, Highly similar to COATOMER ZETA SUBUNIT [Bos taurus]
CCTGTAATCT	1342	81	27	-	323	3.65	V-erb-b2 avian erythroblastic leukemia viral oncogene homolog 3 (alternative products)
AGGTCCTAGC	1343	391	16	-	1448	3.66	Glutathione-S-transferase pi-1
ACTGAAGGCG	1344	68	15	-	266	3.68	Human metargidin precursor mRNA, complete cds
AAGGAAGATG	1345	24	6	-	94	3.68	PROTEASOME COMPONENT C13 PRECURSOR
CCGACGGGCG	1346	60	14	-	237	3.71	Tag matches ribosomal RNA sequence
GCCCCCAATA	1347	428	6	-	1601	3.73	Lectin, galactoside-binding, soluble, 1 (galectin 1)
AGGATGTGGG	1348	49	9	-	193	3.74	Homo sapiens mRNA for KIAA0706 protein, complete cds
GGAGGCCGAG	1349	26	5	-	103	3.75	ESTs, Weakly similar to allograft inflammatory factor-1 [H.sapiens]
ACCCCCCCGC	1350	65	6	-	251	3.76	Jun D proto-oncogene
CTGGCCTGTG	1351	30	6	-	120	3.80	Homo sapiens mRNA for CIRP, complete cds
CTGGCCTGTG	1352	30	6	-	120	3.80	Villin 2 (ezrin)
CTGGCCTGTG	1353	30	6	-	120	3.80	Homo sapiens clone 23565 unknown mRNA, partial cds
CACCCCCAGG	1354	29	7	-	118	3.80	ESTs
CACCCCCAGG	1355	29	7	-	118	3.80	Human Gps2 (GPS2) mRNA, complete cds
GTGAAACTCC	1356	66	16	-	269	3.81	Human 53K isoform of Type II phosphatidylinositol-4-phosphate 5-kinase (PIPK) mRNA, complete cds
GTGAAACTCC	1357	66	16	-	269	3.81	Human mRNA for KIAA0328 gene, partial cds
AGAATTGCTT	1358	50	12	-	201	3.81	Homo sapiens nephrin (NPHS1) mRNA, complete cds
AGAATTGCTT	1359	50	12	-	201	3.81	H.sapiens mRNA for phosphorylase-kinase, beta subunit
ATGGCCTCCT	1360	19	5	-	76	3.84	Human syntaxin mRNA, complete cds
AACTGTCCTT	1361	34	5	-	138	3.84	H.sapiens mRNA for major astrocytic phosphoprotein PEA-15
AAGGAATCGG	1362	34	5	-	136	3.85	PROTEASOME BETA CHAIN PRECURSOR
TCTGTTTATC	1363	29	8	-	119	3.86	Signal recognition particle 14 kD protein
ACTTTTTCAA	1364	704	20	-	2741	3.87	Tag matches mitochondrial sequence
TCTGTAATCC	1365	46	8	-	185	3.87	Tag matches mitochondrial sequence
TCTGTAATCC	1366	46	8	-	185	3.87	Human aryl sulfotransferase mRNA, complete cds
GTGAAAACCC	1367	27	5	-	110	3.90	No match
GGCAGGCACA	1368	24	5	-	97	3.91	H.sapiens mRNA for phenylalkylamine binding protein
GGGGCAGGGC	1369	281	33	-	1138	3.93	ESTs, Weakly similar to EPIDERMAL GROWTH FACTOR PRECURSOR, KIDNEY
GGGGCAGGGC	1370	281	33	-	1138	3.93	Eukaryotic translation initiation factor 5A
GTGAAACTCT	1371	32	8	-	134	3.94	No match
TGGACCAGGC	1372	28	7	-	118	3.95	ESTs, Weakly similar to No definition line found [C.elegans]
CCTATAATCC	1373	109	16	-	452	4.01	Retinoblastoma-like 1 (p107)
CCTATAATCC	1374	109	16	-	452	4.01	Cyclic nucleotide gated channel (photoreceptor), cGMP gated 2 (beta)
CCTATAATCC	1375	109	16	-	452	4.01	Homo sapiens mRNA for KIAA0694 protein, complete cds
AACTGCTTCA	1376	77	12	-	323	4.05	Homo sapiens Arp2/3 protein complex subunit p41-Arc (ARC41) mRNA, complete cds
GGATTGTCTG	1377	55	11	-	233	4.07	Small nuclear ribonucleoprotein polypeptides B and B1
CCTGTAATTC	1378	48	8	-	201	4.07	Homo sapiens mRNA for KIAA0591 protein, partial cds
CTGGGCCTGG	1379	84	7	-	351	4.07	Human HU-K4 mRNA, complete cds
ACCCTTGGCC	1380	551	83	-	2334	4.08	Tag matches mitochondrial sequence
ATGGCGATCT	1381	27	7	-	117	4.09	Ribosomal protein S24
TTGTCTGCCT	1382	39	8	-	166	4.10	ESTs
TGAATCTGGG	1383	35	6	-	150	4.11	SET translocation (myeloid leukemia-associated)
AGCCTTTGTT	1384	57	6	-	240	4.13	Human mRNA for collagen binding protein 2, complete cds
CTTTTCAGCA	1385	29	9	-	129	4.17	Human 14-3-3 epsilon mRNA, complete cds

Table 4, cont.

CCTGGAGTGG	1386	28	5	123	4.17	ESTs
CGGAGACCCT	1387	87	14	380	4.20	Homo sapiens dbpB-like protein mRNA, complete cds
CCCTGGGTTC	1388	1027	93	4414	4.21	Ferritin, light polypeptide
ATTTGAGAAG	1389	643	93	2814	4.23	Tag matches mitochondrial sequence
ACAACTCAAT	1390	61	6	265	4.24	ESTs, Highly similar to BRAIN PROTEIN I3 [Mus musculus]
CTTGATTCCC	1391	45	8	202	4.30	Homo sapiens quiescin (Q6) mRNA, complete cds
GGCTGGTCTC	1392	48	9	216	4.32	ESTs
AGGTGGCAAG	1393	194	45	891	4.36	Tag matches mitochondrial sequence
CTAGCTTTTA	1394	46	10	210	4.36	Tag matches mitochondrial sequence
TCACCGGTCA	1395	143	23	648	4.38	GELSOLIN PRECURSOR, PLASMA
GGCCGCGTTC	1396	110	5	487	4.38	Ribosomal protein S17
GAGAGCTCCC	1397	84	6	280	4.41	Tag matches mitochondrial sequence
GAGAGCTCCC	1398	84	6	280	4.41	EST
GAGAGCTCCC	1399	84	6	280	4.41	ESTs
GAGAGCTCCC	1400	84	6	280	4.41	Homo sapiens clone 24751 unknown mRNA
CCCCGTACAT	1401	122	7	549	4.43	No match
TGGCGTACGG	1402	67	11	314	4.50	Tag matches ribosomal RNA sequence
TCCCCGACAT	1403	97	5	444	4.53	No match
CCTGGCTAAT	1404	32	11	155	4.53	No match
TCACAGCTGT	1405	50	10	238	4.61	B-cell translocation gene 1, anti-proliferative
TCCCATTAAG	1406	119	12	560	4.61	No match
GTGCACTGAG	1407	259	21	1228	4.65	Major histocompatibility complex, class I, C
GTGCACTGAG	1408	259	21	1228	4.65	MHC class I protein HLA-A (HLA-A28.-B40. -Cw3)
GCTTACCTTT	1409	35	6	170	4.68	Homo sapiens calumein (Calu) mRNA, complete cds
CTGGCCCGGA	1410	54	7	284	4.71	Vasodilator-stimulated phosphoprotein
CTGGCCCGGA	1411	54	7	284	4.71	Homo sapiens Sox-like transcriptional factor mRNA, complete cds
GGGCCTGTGC	1412	133	11	647	4.79	Homo sapiens monocarboxylate transporter (MCT3) mRNA, complete cds
GGGCCTGTGC	1413	133	11	647	4.79	ESTs
GCCCCTCCGG	1414	121	18	598	4.79	ESTs, Weakly similar to TRANS-ACTING TRANSCRIPTIONAL PROTEIN ICP0
TTGTGATGTA	1415	21	5	109	4.87	Neurotrophic tyrosine kinase, receptor, type 1
TTGTGATGTA	1416	21	5	109	4.87	Fibroblast growth factor receptor 4
CATCTTCACC	1417	62	5	311	4.97	Ribosomal protein S25
TTGGCCAGGA	1418	100	35	539	5.06	No match
AGAATCACTT	1419	37	5	194	5.09	No match
TTAGCCAGGA	1420	23	8	129	5.22	Human LLGL mRNA, complete cds
GTTGTGGTTA	1421	496	43	2646	5.25	BETA-2-MICROGLOBULIN PRECURSOR
CAAGCATCCC	1422	547	36	2910	5.26	Tag matches mitochondrial sequence
GACATATGTA	1423	38	8	217	5.29	Cytochrome c oxidase subunit VIIb
AGTATCTGGG	1424	83	8	337	5.29	Homo sapiens Arp2/3 protein complex subunit p41-Arc (ARC41) mRNA, complete cds
ACCGCCTGTG	1425	120	19	659	5.35	Human transcriptional activator mRNA, complete cds
CTCTTCGAGA	1426	177	15	983	5.35	Glutathione peroxidase 1
ATGAGCTGAC	1427	104	11	571	5.42	CYSTATIN B
GCCTCTGTCT	1428	38	5	202	5.43	Ribosomal protein, large, P1
AAGGAAGATC	1429	38	6	214	5.43	Human glutathione-S-transferase homolog mRNA, complete cds
AAAACATTCT	1430	308	30	1698	5.45	Tag matches mitochondrial sequence
CTCAGACAGT	1431	64	5	385	5.95	ESTs, Highly similar to 40S RIBOSOMAL PROTEIN S27 [Rattus norvegicus]
CCCAAGCTAG	1432	435	54	2698	6.08	Heat shock 27kD protein 1

Table 4, cont.

CCCAAGCTAG	1433	435	54	2698	6.08	Tag matches ribosomal RNA sequence
TCAATCAAGA	1434	34	8	236	6.67	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, eta polypeptide
TGCAGCGCCT	1435	111	9	762	6.80	H.sapiens mRNA for uridine phosphorylase
TTCACTGTGA	1436	223	7	1557	6.94	Lectin, galactoside-binding, soluble, 3 (galectin 3) (NOTE: redefinition of symbol)
CTGACCTGTG	1437	226	16	1683	7.38	HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-27 ALPHA CHAIN PRECURSOR
GGGGTCAGGG	1438	118	9	862	7.43	Glycogen phosphorylase B (brain form)
GGCTTTAGGG	1439	125	10	1019	8.05	Tag matches mitochondrial sequence
TGGGTGAGCC	1440	304	45	2538	8.21	Cathepsin B
AGGGTGTTTT	1441	78	8	668	8.43	Dual-specificity tyrosine-(Y)-phosphorylation regulated kinase
AGGGTGTTTT	1442	78	8	668	8.43	Tag matches mitochondrial sequence
TGGTGTATGC	1443	93	6	810	8.62	Tag matches mitochondrial sequence
GAGTAGAGAA	1444	50	8	465	9.15	SET translocation (myeloid leukemia-associated)
TGCAGGCCTG	1445	115	11	1165	10.02	TRYPTOPHANYL-TRNA SYNTHETASE
GCGAAACCCT	1446	210	34	2242	10.51	V-erb-b2 avian erythroblastic leukemia viral oncogene homolog 3 (alternative products)
GTGACCACGG	1447	4374	29	47260	10.80	Human N-methyl-D-aspartate receptor 2C subunit precursor (NMDAR2C) mRNA, complete cds
GTGACCACGG	1448	4374	29	47260	10.80	Tag matches ribosomal RNA sequence

Table 5. Transcripts uniformly elevated in cancer tissues

Tag Sequence	SEQ ID NO:	Cancer tissues CC	BC	BrC	LC	M	Normal Tissues NC	NB	NBr	NL	NM	Avg T/N	UniGene Description
ATGTGTAACG	226	93	72	13	5	48	0	0	3	0	0	30	S100 calcium-binding protein A4 (calcium protein, calvasculin, metastasin)
CCCTGCCTTG	227	53	66	120	56	20	21	27	0	8	0	21	Midkine (neurite growth-promoting factor 2)
GTGCGCTGAG	228	85	103	380	23	58	0	30	56	0	8	18	Major histocompatibility complex, class I, C
CTGGCCGCTC	229	26	19	53	16	25	3	1	0	0	5	14	Apoptosis inhibitor 4 (survivin)
GCCCCCCCGT	230	38	40	54	31	29	9	7	3	3	0	12	ESTs
TGGCCCCAGG	231	13	201	8	24	336	0	30	3	3	19	9	Apolipoprotein CI
CCCTGGTGGG	232	16	14	17	16	6	0	0	0	0	3	9	ESTs
AGTGACCGAA	233	5	8	37	8	7	0	1	0	3	0	8	ESTs
CTGCACTTAC	234	52	34	81	64	78	3	12	22	5	30	8	DNA REPLICATION LICENSING FACTOR CDC47 HOMOLOG
CTGGCGAGCG	235	168	137	290	73	178	9	21	64	13	60	8	Human ubiquitin carrier protein (E2-EPF) mRNA, complete cds
TTGCCGCTGC	236	4	10	12	19	7	0	1	0	0	0	7	ESTs
TGCGCTGGCC	237	22	63	74	28	14	6	16	6	8	0	7	No match
CTCCTGGAAC	238	20	10	26	16	16	3	4	0	8	5	6	ESTs, Highly similar to MYO-INOSITOL-1-PHOSPHATE SYNTHASE [Arabidopsis thaliana]
CGCCCGTCGT	239	4	151	30	9	30	0	13	6	0	5	6	No match
TTGCCCCCGT	240	10	61	15	19	23	0	22	6	5	0	6	AXL receptor tyrosine kinase
TTGCTAAAGG	241	8	8	16	16	22	3	0	3	8	0	6	ESTs, Weakly similar to KIAA0005 [H.sapiens]
AGCCACGTTG	242	13	6	11	11	6	0	0	0	0	3	6	Acid phosphatase 1, soluble
CCTGGGCACT	243	14	6	23	22	8	3	1	3	3	0	6	ESTs, Highly similar to transcription factor ARF6 chain B [M.musculus]
GGGCTCACCT	244	23	13	52	16	17	3	4	6	3	5	6	Homo sapiens clone 24767 mRNA sequence / ESTs, Weakly similar to colt [D.melanogaster]
CTTACAGCCA	245	11	6	19	12	6	0	0	3	0	3	6	ESTs
AGGGCCCTCA	246	14	6	15	5	4	0	3	0	0	0	6	Homo sapiens mRNA, complete cds
GGGTAATGTG	247	7	13	5	11	12	0	1	0	0	5	5	ESTs, Moderately similar to unknown [M.musculus]
CTGACAGCCC	248	4	5	17	7	9	0	1	0	0	3	5	Human mRNA for HsMcm6, complete cds
TGACCTCCAG	249	7	14	15	12	11	0	6	3	3	0	5	ESTs, Weakly similar to No definition line found [C.elegans] / ESTs
AAACCTCTTC	250	10	5	12	11	8	0	1	3	0	3	5	ESTs, Highly similar to G2/MITOTIC-SPECIFIC CYCLIN B2 [Mesocricetus auratus]
TCATTGCACT	251	7	13	5	4	9	3	1	0	0	0	5	ESTs, Highly similar to HYPOTHETICAL 16.3 KD PROTEIN [Saccharomyces cerevisiae]
CCCCCTCCGG	252	31	14	73	38	58	15	3	8	19	11	5	Small nuclear ribonucleoprotein polypeptide N / B and B1
GTAGGGGCCT	253	11	14	11	19	18	3	6	0	3	8	4	ESTs
GAACCCAAAG	254	7	8	12	8	10	0	0	3	3	3	4	Plasminogen / PEPTIDYL-PROLYL CIS-TRANS ISOMERASE A
TGTGAGCCTC	255	5	11	11	7	7	0	3	0	0	3	4	Cyclin F
ATCTCTGGAG	256	7	3	9	8	7	0	0	0	0	3	4	ESTs
AAAGTGCATC	257	10	19	11	4	7	0	9	0	0	3	4	No match
GCCTTGGGTG	258	7	8	4	9	10	3	3	0	0	0	4	Leukemia inhibitory factor (cholinergic differentiation factor)
ACCTCACTCT	259	9	3	12	16	9	0	0	6	3	3	4	ESTs
TAAAGACTTG	260	9	13	24	12	38	3	1	11	5	11	4	Adenylate kinase 2 (adk2)
TCGGCGCCGG	261	15	16	21	14	6	6	3	8	3	0	4	SET translocation (myeloid leukemia-associated)
AACCTCGAGT	262	6	10	7	8	11	0	4	0	3	3	4	ESTs, Moderately similar to putative [M.musculus]
GTTTACCCGC	263	6	3	4	7	4	0	0	0	0	0	3	No match
GCCTCTGCCT	264	4	5	5	5	6	0	0	0	0	3	3	ESTs
CCTGGGTCCT	265	4	10	8	5	7	0	4	3	0	3	3	ESTs

Table 6. Transcripts expressed in Colon Cancer Cells (>500 copies per cell)

Tag	SEQ ID NO:	Copies/cell	Unigene Description
CCCATCGTCC	1449	2672	Tag matches mitochondrial sequence
TGTGTTGAGA	1450	1672	Translation elongation factor 1-alpha-1
GGATTTGGCC	1451	1663	Ribosomal protein, large P2 / Ribosomal protein S26 / Human mRNA for PIG-B, complete cds
CCCGTCCGGA	1452	1559	60S RIBOSOMAL PROTEIN L13
ATGGCTGGTA	1453	1555	40S RIBOSOMAL PROTEIN S2
GTGAAACCCC	1454	1482	Multiple matches
CCTCCAGCTA	1455	1468	Keratin 8
TTGGTCCTCT	1456	1453	60S RIBOSOMAL PROTEIN L41
TGATTTCACT	1457	1434	EST / Tag matches mitochondrial sequence
CCTGTAATCC	1458	1372	Multiple matches
ACTTTTTCAA	1459	1367	Tag matches mitochondrial sequence
AAAAAAAAAA	1460	1357	Multiple matches
GAGGGAGTTT	1461	1290	Ribosomal protein L27a
GCCGAGGAAG	1462	1141	Human mRNA for ribosomal protein S12
CACCTAATTG	1463	1137	Tag matches mitochondrial sequence
CGCCGCCGGC	1464	1098	Human ribosomal protein L35 mRNA, complete cds
GGGGAAATCG	1465	1092	THYMOSIN BETA-10
GAAAAATGGT	1466	1056	Laminin receptor (2H5 epitope)
GGGCTGGGGT	1467	1028	H.sapiens mRNA for ribosomal protein L29 / Homo sapiens sperm acrosomal protein mRNA
GCCGGGTGGG	1468	986	Basigin
AGCCCTACAA	1469	945	Tag matches mitochondrial sequence
CTGGGTTAAT	1470	943	40S RIBOSOMAL PROTEIN S19
CAAACCATCC	1471	927	Keratin 18
TGCACGTTTT	1472	916	Human mRNA for antileukoprotease (ALP) from cervix uterus
AGGCTACGGA	1473	905	60S RIBOSOMAL PROTEIN L13A
GCAGCCATCC	1474	861	Ribosomal protein L28
TTCAATAAAA	1475	851	Ribosomal protein, large, P1 / TRANSCOBALAMIN I PRECURSOR
CTAAGACTTC	1476	833	Tag matches mitochondrial sequence
TGGTGTTGAG	1477	830	Human DNA sequence from clone 1033B10 on chromosome 6p21.2-21.31
TACCATCAAT	1478	828	Glyceraldehyde-3-phosphate dehydrogenase
TTCATACACC	1479	814	Tag matches mitochondrial sequence
CCACTGCACT	1480	800	Multiple matches
ACTAACACCC	1481	795	Tag matches mitochondrial sequence
AAGGTGGAGG	1482	794	60S RIBOSOMAL PROTEIN L18A
AGCACCTCCA	1483	787	Eukaryotic translation elongation factor 2
CACAAACGGT	1484	761	40S RIBOSOMAL PROTEIN S27
AGGAAAGCTG	1485	732	ESTs, Highly similar to 60S RIBOSOMAL PROTEIN L36 [Rattus norvegicus]
GTGAAACCCT	1486	729	Multiple matches
AATCCTGTGG	1487	711	Ribosomal protein L8
TTGGGGTTTC	1488	698	Ferritin heavy chain
AAGACAGTGG	1489	696	Ribosomal protein L37a
ATTTGAGAAG	1490	680	Tag matches mitochondrial sequence
GCCGTGTCCG	1491	679	Human ribosomal protein S6 mRNA, complete cds

Table 6, cont.

CGCCGGAACA	1492	678	Ribosomal protein L4
TCTCCATACC	1493	661	Tag matches mitochondrial sequence
ACATCATCGA	1494	661	Ribosomal protein L12
AACGCGGCCA	1495	644	Macrophage migration inhibitory factor
AGGGCTTCCA	1496	643	UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX SUBUNIT VI REQUIRING PROTEIN
CCGTCCAAGG	1497	631	Ribosomal protein S16
CGCTGGTTCC	1498	626	Homo sapiens ribosomal protein L11 mRNA, complete cds
CTCAACATCT	1499	615	Ribosomal protein, large, P0
ACTCCAAAAA	1500	608	H.sapiens mRNA for transmembrane protein mp24 / Human insulinoma rig-analog mRNA encoding DNA-binding protein
CCTAGCTGGA	1501	606	PEPTIDYL-PROLYL CIS-TRANS ISOMERASE A
GTGAAGGCAG	1502	596	Ribosomal protein S3A
AGCTCTCCCT	1503	551	60S RIBOSOMAL PROTEIN L23
TAGGTTGTCT	1504	537	TRANSLATIONALLY CONTROLLED TUMOR PROTEIN
GGACCACTGA	1505	522	Ribosomal protein L3
AAGGAGATGG	1506	521	Ribosomal protein L31
AACTAAAAAA	1507	510	Ubiquitin A-52 residue ribosomal protein fusion product 1
GGCTGGGGGC	1508	507	Human profilin mRNA, complete cds
CCAGAACAGA	1509	503	Deoxythymidylate kinase / 60S RIBOSOMAL PROTEIN L30

Table 7. Expressed transcripts (>500 copies per cell)

Tag Sequence	SEQ ID NO:	Copies/Cell	Description
CCCATCGTCC	1508	3022	Tag matches mitochondrial sequence
GTGACCACGG	1509	2435	Tag matches ribosomal RNA sequence / Human N-methyl-D-aspartate receptor 2C subunit precursor (NMDAR2C) mRNA
TGTGTTGAGA	1510	1557	Translation elongation factor 1-alpha-1
GTGAAACCCC	1511	1466	Multiple matches
CCTGTAATCC	1512	1403	Multiple matches
CTAAGACTTC	1513	1349	Tag matches mitochondrial sequence
CACCTAATTG	1514	1333	Tag matches mitochondrial sequence
CCCGTCCGGA	1515	1282	60S RIBOSOMAL PROTEIN L13
TTGGTCCTCT	1516	1238	60S RIBOSOMAL PROTEIN L41
ATGGCTGGTA	1517	1126	40S RIBOSOMAL PROTEIN S2
TTGGGGTTTC	1518	1099	Ferritin heavy chain
CCACTGCACT	1519	964	Multiple matches
TGATTTCACT	1520	942	Tag matches mitochondrial sequence / EST
ACTTTTTCAA	1521	899	Tag matches mitochondrial sequence
GCAGCCATCC	1522	886	Ribosomal protein L28
TACCATCAAT	1523	874	Glyceraldehyde-3-phosphate dehydrogenase
GGATTTGGCC	1524	854	Ribosomal protein, large P2 / Ribosomal protein S26 / Human mRNA for PIG-B
CCCTGGGTTC	1525	844	Ferritin, light polypeptide
GCCGAGGAAG	1526	836	Human mRNA for ribosomal protein S12
AGGCTACGGA	1527	820	60S RIBOSOMAL PROTEIN L13A
CGCCGCCGGC	1528	805	Human ribosomal protein L35 mRNA, complete cds
TTCATACACC	1529	804	Tag matches mitochondrial sequence
AGCCCTACAA	1530	801	Tag matches mitochondrial sequence
CACAAACGGT	1531	799	40S RIBOSOMAL PROTEIN S27
AAGGTGGAGG	1532	786	60S RIBOSOMAL PROTEIN L18A
CTTCCTTGCC	1533	777	Keratin 17
TGGTGTTGAG	1534	770	Human DNA sequence from clone 1033B10 on chromosome 6p21.2-21.31
GTGAAACCCT	1535	728	Multiple matches
GGGGAAATCG	1536	724	THYMOSIN BETA-10
AGCACCTCCA	1537	718	Eukaryotic translation elongation factor 2
CCTCCAGCTA	1538	711	Keratin 8
AAGACAGTGG	1539	699	Ribosomal protein L37a
CTGGGTTAAT	1540	699	40S RIBOSOMAL PROTEIN S19
ATTTGAGAAG	1541	689	Tag matches mitochondrial sequence
GCCGGGTGGG	1542	687	Basigin
GGGCTGGGGT	1543	683	H.sapiens mRNA for ribosomal protein L29 / Homo sapiens sperm acrosomal protein mRNA
AGGGCTTCCA	1544	663	UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX SUBUNIT VI REQUIRING PROTEIN
AAAAAAAAAA	1545	650	Multiple matches
GAGGGAGTTT	1546	648	Ribosomal protein L27a
GCGACCGTCA	1547	637	Aldolase A
ACTAACACCC	1548	631	Tag matches mitochondrial sequence
CGCCGGAACA	1549	616	Ribosomal protein L4
TGGGCAAAGC	1550	592	Translation elongation factor 1 gamma
TGCACGTTTT	1551	586	Human mRNA for antileukoprotease (ALP) from cervix uterus

Table 7, cont.

AATCCTGTGG	1552	569	Ribosomal protein L8
CAAGCATCCC	1553	565	Tag matches mitochondrial sequence
CCGTCCAAGG	1554	559	Ribosomal protein S16
TAGGTTGTCT	1555	551	TRANSLATIONALLY CONTROLLED TUMOR PROTEIN
GCCGTGTCCG	1556	540	Human ribosomal protein S6 mRNA, complete cds
GCTTTATTTG	1557	540	Human mRNA fragment encoding cytoplasmic actin
CTAGCCTCAC	1558	539	Actin, gamma 1
CCTAGCTGGA	1559	537	PEPTIDYL-PROLYL CIS-TRANS ISOMERASE A
GCCCCTGCTG	1560	534	Keratin 5 (epidermolysis bullosa simplex, Dowling-Meara/Kobner/Weber-Cockayne types)
ACCCTTGGCC	1561	526	Tag matches mitochondrial sequence
AGGAAAGCTG	1562	513	ESTs, Highly similar to 60S RIBOSOMAL PROTEIN L36 [Rattus norvegicus]

CLAIMS

1. A method of identifying a cell as either a colon epithelial cell, a brain cell, a keratinocyte, a breast epithelial cell, a lung epithelial cell, a melanocyte, a prostate cell, or a kidney epithelial cell, comprising the step of:

determining expression in a test cell of a gene product of at least one gene comprising a sequence selected from at least one of the following groups:

(a) the sequences shown in SEQ ID NOS:2, 5-18, 20-84, and 85;

(b) the sequences shown in SEQ ID NOS:87-96, 98, 100-103, 105, 107-110, 112-129, and 131-150, and 151;

(c) the sequences shown in SEQ ID NOS:152-154, and 155;

(d) the sequences shown in SEQ ID NOS:156-159, and 160;

(e) the sequences shown in SEQ ID NOS:161-166, and 167;

(f) the sequences shown in SEQ ID NOS:168, 170, 172-177, 179-188, 190-207, and 208;

(g) the sequences shown in SEQ ID NOS:209 and 210; and

(h) the sequences shown in SEQ ID NOS:211-224 and 225,

wherein expression of a gene product of at least one gene comprising a sequence shown in (a) identifies the test cell as a colon epithelial cell;

wherein expression of a gene product of at least one gene comprising a sequence shown in (b) identifies the test cell as a brain cell;

wherein expression of a gene product of at least one gene comprising a sequence shown in (c) identifies the test cell as a keratinocyte;

wherein expression of a gene product of at least one gene comprising a sequence shown in (d) identifies the test cell as a breast epithelial cell;

wherein expression of a gene product of at least one gene comprising a sequence shown in (e) identifies the test cell as a lung epithelial cell;

wherein expression of a gene product of at least one gene comprising a sequence shown in (f) identifies the test cell as a melanocyte;

wherein expression of a gene product of at least one gene comprising a sequence shown in (g) identifies the test cell as a prostate cell; and

wherein expression of a gene product of at least one gene comprising a sequence shown in (h) identifies the test cell as a kidney epithelial cell.

2. The method of claim 1 wherein expression of gene products of at least two of said genes is determined.

3. The method of claim 1 wherein expression of gene products of at least five of said genes is determined.

4. The method of claim 1 wherein the gene product is protein.

5. The method of claim 1 wherein the gene product is RNA.

6. The method of claim 5 wherein expression is determined using at least one oligonucleotide probe.

7. The method of claim 5 wherein expression is determined using at least two oligonucleotide probes.

8. The method of claim 6 wherein the at least one oligonucleotide probe is immobilized on a solid support.

9. The method of claim 8 wherein the at least one oligonucleotide probe is in an array.

10. The method of claim 1 wherein the cell to be identified is a cancer cell.

11. An isolated polynucleotide comprising a sequence selected from the group consisting of SEQ ID NOS:2, 5, 6, 8, 10, 12, 13, 15, 17, 18, 21, 24-26, 28, 30, 31, 34-36, 38, 40, 47-51, 53-57, 59-62, 65-69, 71-76, 78, 80-84, 98, 103, 113, 115, 122, 129, 132, 134, 135, 140, 144, 149, 150, 153-168, 174-176, 182, 185, 186, 188, 190, 200, 201, 205-213, 216-224, 237, 239, 257, 263, 485, 487, 495, 499, 514, 586, 686, 751, 835, 844, 878, 910, 925, 932, 951, 1000, 1005, 1070, 1122, 1130, 1170, 1173, 1187, 1189, 1200, 1213, 1220, 1237, 1257, 1264, 1273, 1293, 1300, 1320, 1367, 1371, 1401, 1403, 1404, 1406, 1418, and 1419.

12. A solid support comprising at least one polynucleotide comprising a sequence selected from at least one of the following groups:

(a) the sequences shown in SEQ ID NOS:2, 5, 6, 8, 10, 12, 13, 15, 17, 18, 21, 24-26, 28, 30, 31, 34-36, 38, 40, 47-51, 53-57, 59-62, 65-69, 71-76, 78, 80-83, and 84;

(b) the sequences shown in SEQ ID NOS:98, 103, 113, 115, 122, 129, 132,

134, 135, 140, 144, 149, and 150;

(c) the sequences shown in SEQ ID NOS:153-154 and 155;

(d) the sequences shown in SEQ ID NOS:156-157 and 160;

(e) the sequences shown in SEQ ID NOS:161-166 and 167;

(f) the sequences shown in SEQ ID NOS:168, 174-176, 182, 185, 186, 188, 190, 200, 201, 205-207 and 208;

(g) the sequences shown in SEQ ID NOS:209 and 210;

(h) the sequences shown in SEQ ID NOS:211-213, 216-223, and 224;

(i) the sequences shown in SEQ ID NOS:237, 239, 257, and 263; or

(j) the sequences shown in SEQ ID NOS:485, 487, 495, 499, 514, 586, 686, 751, 835, 844, 878, 910, 925, 932, 951, 1000, 1005, 1070, 1122, 1130, 1170, 1173, 1187, 1189, 1200, 1213, 1220, 1237, 1257, 1264, 1273, 1293, 1300, 1320, 1367, 1371, 1401, 1403, 1404, 1406, 1418, and 1419.

13. The solid support of claim 12 wherein:

if the at least one polynucleotide comprises a sequence selected from (a), then the solid support further comprises a polynucleotide comprising a sequence selected from the group consisting of the sequences shown in SEQ ID NOS:1, 3, 4, 7, 9, 11, 14, 16, 19, 20, 22, 23, 27, 29, 32, 33, 37, 39, 41-46, 52, 58, 63, 64, 70, 77, 79, and 85;

if the at least one polynucleotide comprises a sequence selected from (b), then the solid support further comprises a polynucleotide comprising a sequence selected from the group consisting of the sequences shown in SEQ ID NOS:86-97, 99-102, 104-112, 114, 116-121, 123-128, 130, 131, 133, 136-139, 141-143, 145-148, and 151;

if the at least one polynucleotide comprises a sequence selected from (c), then the solid support further comprises a polynucleotide comprising the sequence shown in SEQ ID NO:152;

if the at least one polynucleotide comprises a sequence selected from (f), then the solid support further comprises a polynucleotide comprising a sequence selected from the group consisting of the sequences shown in SEQ ID NOS:169-173, 177-181, 183, 184, 187, 189, 191-199, 202, 203, and 204;

if the at least one polynucleotide comprises a sequence selected from (h), then

the solid support further comprises a polynucleotide comprising a sequence selected from the group consisting of the sequences shown in SEQ ID NOS:214, 215, and 225;

if the at least one polynucleotide comprises a sequence selected from (i), then the solid support further comprises a polynucleotide comprising a sequence selected from the group consisting of the sequences shown in SEQ ID NOS:226-236, 238, 240-256, 258-262, 264, and 265; and

if the at least one polynucleotide comprises a sequence selected from (j), then the solid support further comprises a polynucleotide comprising a sequence selected from the group consisting of the sequences shown in SEQ ID NOS:266-484, 486, 488-494, 496-498, 500-513, 515-585, 587-685, 687-750, 752-834, 836-843, 845-877, 879-909, 911-924, 926-931, 933-950, 952-999, 1001-1004, 1006-1069, 1071-1121, 1123-1129, 1131-1169, 1171, 1171, 1174-1186, 1188, 1190-1199, 1201-1212, 1214-1219, 1221-1236, 1238-1256, 1258-1263, 1265-1272, 1274-1292, 1294-1299, 1301-1319, 1321-1366, 1368-1370, 1372-1400, 1402, 1405, 1407-1416, and 1417.

14. The solid support of claim 12, wherein:

if the at least one polynucleotide comprises a sequence selected from (a), then the at least one polynucleotide further comprises a sequence selected from the group consisting of the sequences shown in SEQ ID NOS:1, 3, 4, 7, 9, 11, 14, 16, 19, 20, 22, 23, 27, 29, 32, 33, 37, 39, 41-46, 52, 58, 63, 64, 70, 77, 79, and 85;

if the at least one polynucleotide comprises a sequence selected from (b), then the at least one polynucleotide further comprises a sequence selected from the group consisting of the sequences shown in SEQ ID NOS:86-97, 99-102, 104-112, 114, 116-121, 123-128, 130, 131, 133, 136-139, 141-143, 145-148, and 151;

if the at least one polynucleotide comprising a sequence selected from (c), then the at least one polynucleotide further comprises SEQ ID NO:152;

if the at least one polynucleotide comprises a sequence selected from (f), then the at least one polynucleotide further comprises a sequence selected from the group consisting of the sequences shown in SEQ ID NOS:169-173, 177-181, 183, 184, 187, 189, 191-199, 202, 203, and 204;

if the at least one polynucleotide comprises a sequence selected from (h), then

the at least one polynucleotide further comprises a sequence selected from the group consisting of the sequences shown in SEQ ID NOS:214, 215, and 225;

if the at least one polynucleotide comprises a sequence selected from (i), then the at least one polynucleotide further comprises a sequence selected from the group consisting of the sequences shown in SEQ ID NOS:226-236, 238, 240-256, 258-262, 264, and 265; and

if the at least one polynucleotide comprises a sequence selected from (j), then the at least one polynucleotide further comprises a sequence selected from the group consisting of the sequences shown in SEQ ID NOS:266-484, 486, 488-494, 496-498, 500-513, 515-585, 587-685, 687-750, 752-834, 836-843, 845-877, 879-909, 911-924, 926-931, 933-950, 952-999, 1001-1004, 1006-1069, 1071-1121, 1123-1129, 1131-1169, 1171, 1171, 1174-1186, 1188, 1190-1199, 1201-1212, 1214-1219, 1221-1236, 1238-1256, 1258-1263, 1265-1272, 1274-1292, 1294-1299, 1301-1319, 1321-1366, 1368-1370, 1372-1400, 1402, 1405, 1407-1416, and 1417.

15. The solid support of claim 12 wherein the at least one polynucleotide is in an array.

16. A method of identifying a test cell as a cancer cell, comprising the step of:

determining expression in a test cell of a gene product of at least one gene comprising a sequence selected from the group consisting of SEQ ID NOS:228, 230-257, 259-260, and 262-265, wherein an increase in said expression of at least two-fold relative to expression of the at least one gene in a normal cell identifies the test cell as a cancer cell.

17. The method of claim 16 wherein expression of gene products of at least two of said genes is determined.

18. The method of claim 16 wherein expression of gene products of at least five of said genes is determined.

19. The method of claim 16 wherein the gene product is protein.

20. The method of claim 16 wherein the gene product is RNA.

21. The method of claim 20 wherein expression is determined using at least one oligonucleotide probe.

22. The method of claim 21 wherein expression is determined using at least two

oligonucleotide probes.

23. The method of claim 21 wherein the at least one oligonucleotide probe is immobilized on a solid support.

24. The method of claim 23 wherein the at least one oligonucleotide probe is in an array.

25. The method of claim 16 wherein the test cell is selected from the group consisting of a colon epithelial cell, a breast epithelial cell, a lung epithelial cell, a melanocyte, and a brain cell.

26. The method of claim 16 wherein the normal cell and the test cell are selected from a single cell type.

27. A method of reducing expression of a cancer-specific gene in a human cell, comprising the step of:

administering to the cell a reagent which specifically binds to an expression product of a cancer-specific gene comprising a sequence selected from the group consisting of SEQ ID NOS:228, 230-257, 259-260, and 262-265, whereby expression of the cancer-specific gene is reduced relative to expression of the cancer-specific gene in the absence of the reagent.

28. The method of claim 27 wherein the reagent is an antisense oligonucleotide.

29. The method of claim 27 wherein the reagent is an antibody.

30. A method for comparing expression of a gene in a test sample to expression of a gene in a standard sample, comprising the steps of:

determining a first ratio and a second ratio, wherein the first ratio is an amount of an expression product of a test gene in a test sample to an amount of an expression product of at least one gene comprising a sequence selected from the group consisting of SEQ ID NOS:266-375, 377-652, 654-796, and 798-1448 in the test sample, and wherein the second ratio is an amount of an expression product of the test gene in a standard sample to an amount of an expression product of the at least one gene in the standard sample; and

comparing the first and second ratios, wherein a difference between the first and second ratios indicates a difference in the amount of the expression product of the test

gene in the test sample.

31. The method of claim 30 wherein the at least one gene comprises a sequence selected from the group consisting of SEQ ID NOS:282, 288, 300, 302, 308, 320, 323, 363, 368, 379, 381, 444, 453, 518, 531, 535, 538, 542, 579, 580, 594, 600, 604, 617, 626, 641, 650, 717, 728, 776, 777, 794, 818, 822, 842, 885, 887, 899, 900, 902, 904, 914, 930, 960, 964, 1001, 1015, 1020, 1027, 1035, 1090, 1113, 1119, 1146, 1151, 1163, 1233, 1235, 1252, 1255, 1270, 1340, 1345, 1356, 1359, 1360, 1362, 1385, 1415, and 1441.

32. The method of claim 30 wherein expression is determined using at least one oligonucleotide probe.

33. The method of claim 32 wherein the at least one oligonucleotide probe is immobilized on a solid support.

34. The method of claim 33 wherein the at least one oligonucleotide probe is in an array.

35. The method of claim 30 wherein the test sample is a cancer cell and the standard sample is a normal cell.

36. The method of claim 35 wherein the cancer cell is selected from the group consisting of a colon cancer cell, a breast cancer cell, a lung cancer cell, a melanoma cell, and a brain cancer cell.

37. The method of claim 30 wherein the test sample has been treated with a test compound and the standard sample has not been treated with the test compound.

38. The method of claim 37 wherein the test sample is a cancer cell and wherein the standard sample is a normal cell.

39. The method of claim 30 wherein the test sample and the standard sample are obtained from the same cell type.

40. A method of screening candidate anti-cancer drugs, comprising the steps of:

contacting a cancer cell with a test compound; and

measuring expression in the cancer cell of a gene product of at least one gene comprising a sequence selected from the group consisting of SEQ ID NOS: 228, 230-257, 259, 260, 262-263, and 265, wherein a decrease in expression of the gene product in the presence of a test compound relative to expression of the gene product in the absence of the

test compound identifies the test compound as a potential anti-cancer drug.

41. The method of claim 40 wherein the cancer cell is selected from the group consisting of a colon cancer cell, a breast cancer cell, a lung cancer cell, a melanoma cell, and a brain cancer cell.

42. The method of claim 40 in which expression of gene products of at least two of said genes is measured.

43. The method of claim 40 in which expression of gene products of at least five of said genes is measured.

44. The method of claim 40 wherein the gene product is protein.

45. The method of claim 40 wherein the gene product is RNA.

46. The method of claim 45 wherein expression of the at least one gene product is measured using at least one oligonucleotide probe.

47. The method of claim 46 wherein the at least one oligonucleotide probe is immobilized on a solid support.

48. The method of claim 47 wherein the at least one oligonucleotide probe is in an array.

49. The method of claim 46 wherein the at least one oligonucleotide probe comprises a sequence selected from the group consisting of SEQ ID NOS:237, 239, 257, and 263.

50. A method of screening test compounds for the ability to increase an organ or cell function, comprising the step of:

contacting a cell selected from the group consisting of a colon epithelial cell, a brain cell, a keratinocyte, a breast epithelial cell, a lung epithelial cell, a melanocyte, a prostate cell, and a kidney cell with a test compound; and

measuring expression in the cell of a gene product of at least one gene comprising a sequence selected from at least one of the following groups:

(a) the sequences shown in SEQ ID NOS:2, 5-18, 20-84, and 85;

(b) the sequences shown in SEQ ID NOS:87-96, 98, 100-103, 105, 107-110, 112-129, 131-150, and 151;

(c) the sequences shown in SEQ ID NOS:152-154, and 155;

(d) the sequences shown in SEQ ID NOS:156-159 and 160;

(e) the sequences shown in SEQ ID NOS:161-166 and 167;

(f) the sequences shown in SEQ ID NOS:168, 170, 172-177, 179-188, 190-207, and 208;

(g) the sequences shown in SEQ ID NOS:209 and 210; and

(h) the sequences shown in SEQ ID NOS:211-224 and 225,

wherein an increase in expression of a gene product of at least one gene comprising a sequence selected from (a) identifies the test compound as a potential drug for increasing a function of a colon cell;

wherein an increase in expression of a gene product of at least one gene comprising a sequence selected from (b) identifies the test compound as a potential drug for increasing a function of a brain cell;

wherein an increase in expression of a gene product of at least one gene comprising a sequence selected from (c) identifies the test compound as a potential drug for increasing a function of a skin cell;

wherein an increase in expression of a gene product of at least one gene comprising a sequence selected from (d) identifies the test compound as a potential drug for increasing a function of a breast cell;

wherein an increase in expression of a gene product of at least one gene comprising a sequence selected from (e) identifies the test compound as a potential drug for increasing a function of a lung cell;

wherein an increase in expression of a gene product of at least one gene comprising a sequence selected from (f) identifies the test compound as a potential drug for increasing a function of a melanocyte;

wherein an increase in expression of a gene product of at least one gene comprising a sequence selected from (g) identifies the test compound as a potential drug for increasing a function of a prostate cell; and

wherein an increase in expression of a gene product of at least one gene comprising a sequence selected from (h) identifies the test compound as a potential drug for increasing a function of a kidney cell.

51. The method of claim 50 wherein expression of gene products of at least two of said genes is determined.

52. The method of claim 50 wherein expression of gene products of at least five of said genes is determined.

53. The method of claim 50 wherein the gene product is protein.

54. The method of claim 50 wherein the gene product is RNA.

55. The method of claim 54 wherein expression is determined using at least one oligonucleotide probe.

56. The method of claim 54 wherein expression is determined using at least two oligonucleotide probes.

57. The method of claim 55 wherein the at least one oligonucleotide probe is immobilized on a solid support.

58. The method of claim 57 wherein the at least one oligonucleotide probe is in an array.

59. A method to restore function to a diseased tissue or cell comprising the step of:

delivering a gene to a diseased cell selected from the group consisting of a colon epithelial cell, a brain cell, a keratinocyte, a breast epithelial cell, a lung epithelial cell, a melanocyte, a prostate cell, and a kidney cell, wherein the gene comprises a nucleotide sequence selected from at least one of the following groups :

(a) the sequences shown in SEQ ID NOS:2, 5-18, 20-84, and 85;

(b) the sequences shown in SEQ ID NOS:87-96, 98, 100-103, 105, 107-110, 112-129, 131-150, and 151;

(c) the sequences shown in SEQ ID NOS:152-154, and 155;

(d) the sequences shown in SEQ ID NOS:156-159 and 160;

(e) the sequences shown in SEQ ID NOS:161-166 and 167;

(f) the sequences shown in SEQ ID NOS:168, 170, 172-177, 179-188, 190-207, and 208;

(g) the sequences shown in SEQ ID NOS:209 and 210; and

(h) the sequences shown in SEQ ID NOS:211-224 and 225,

wherein expression of the gene in the diseased cell is less than expression of the gene in a corresponding cell which is normal,

wherein if the diseased cell is a colon epithelial cell, then the nucleotide sequence is selected from (a);

wherein if the diseased cell is a brain cell, then the nucleotide sequence is selected from (b);

wherein if the diseased cell is a keratinocyte, then the nucleotide sequence is selected from (c);

wherein if the diseased cell is a breast epithelial cell, then the nucleotide sequence is selected from (d);

wherein if the diseased cell is a lung epithelial cell, then the nucleotide sequence is selected from (e);

wherein if the diseased cell is a melanocyte, then the nucleotide sequence is selected from (f);

wherein if the diseased cell is a prostate cell, then the nucleotide sequence is selected from (g); and

wherein if the diseased cell is a kidney cell, then the nucleotide sequence is selected from (h).

60. The method of claim 59 wherein the diseased cell fails to express the gene in the diseased state.

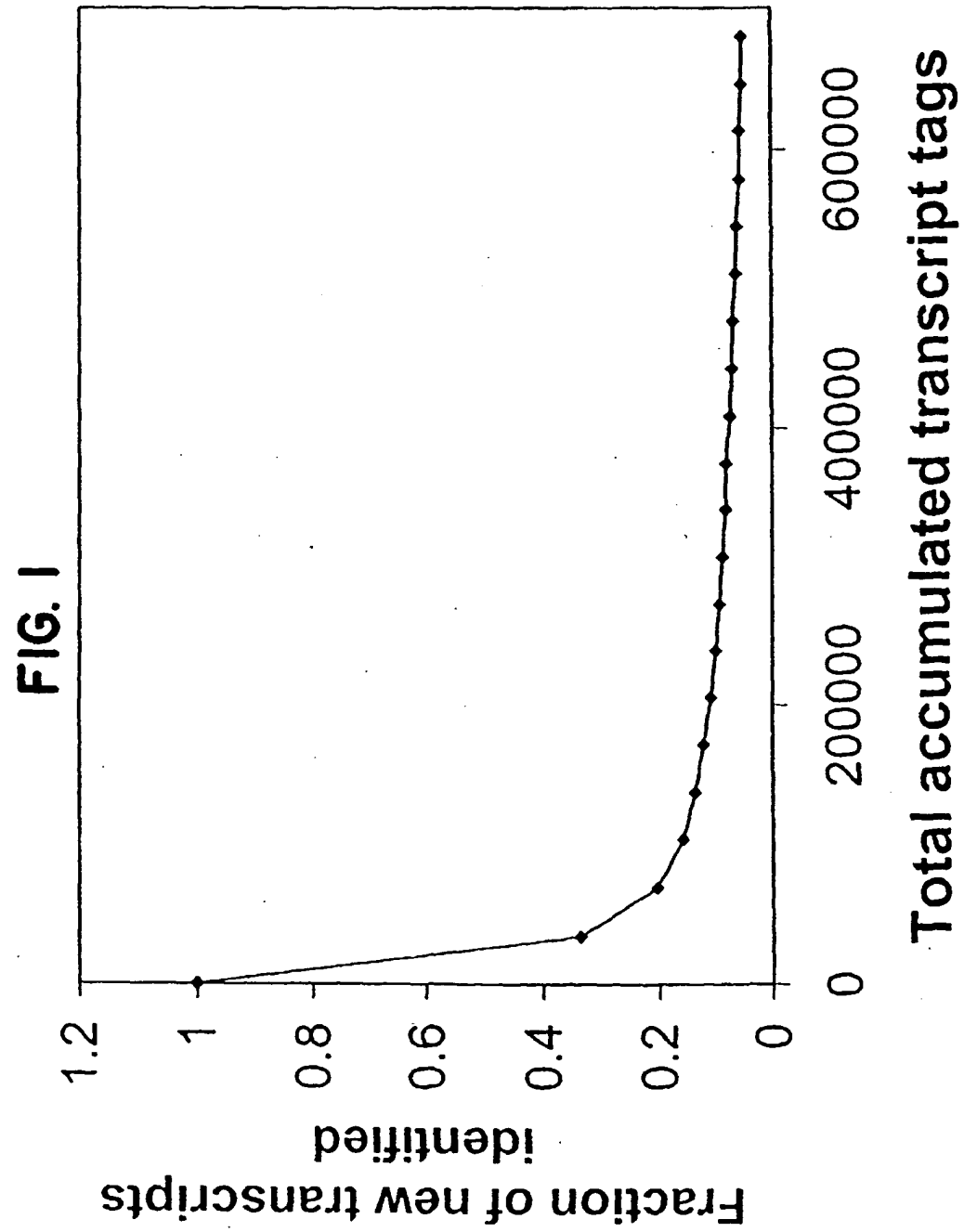
Fraction of new transcripts identified
1.2
1
0.8
0.6
0.4
0.2
0
0
200000
400000
600000
Total accumulated transcript tags

FIG. 1

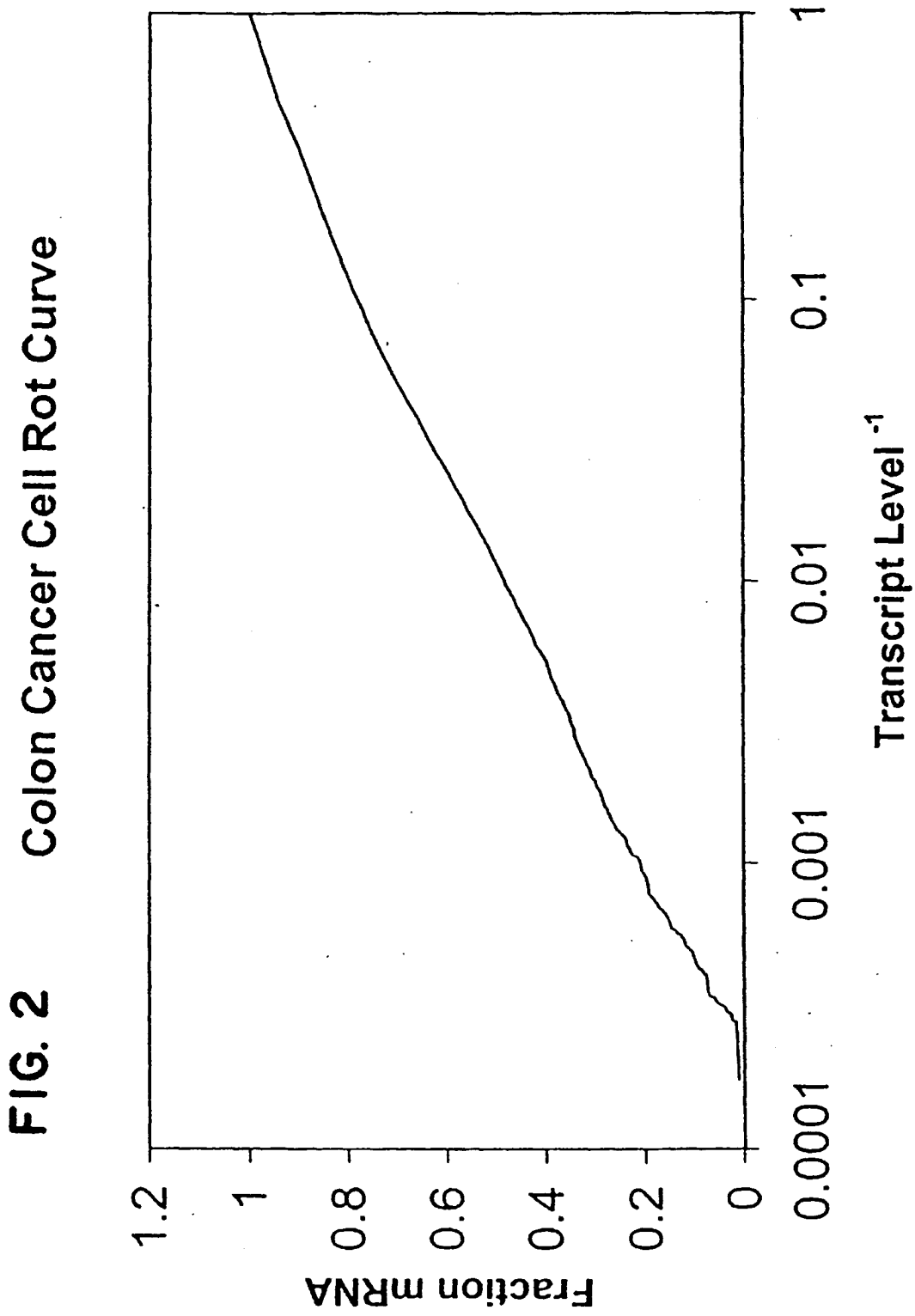
Colon Cancer Cell Rot Curve
Fraction mRNA
1.2
1
0.8
0.6
0.4
0.2
0
0.0001
0.001
0.01
0.1
1
Transcript Level -1

FIG. 2

FIG. 3A

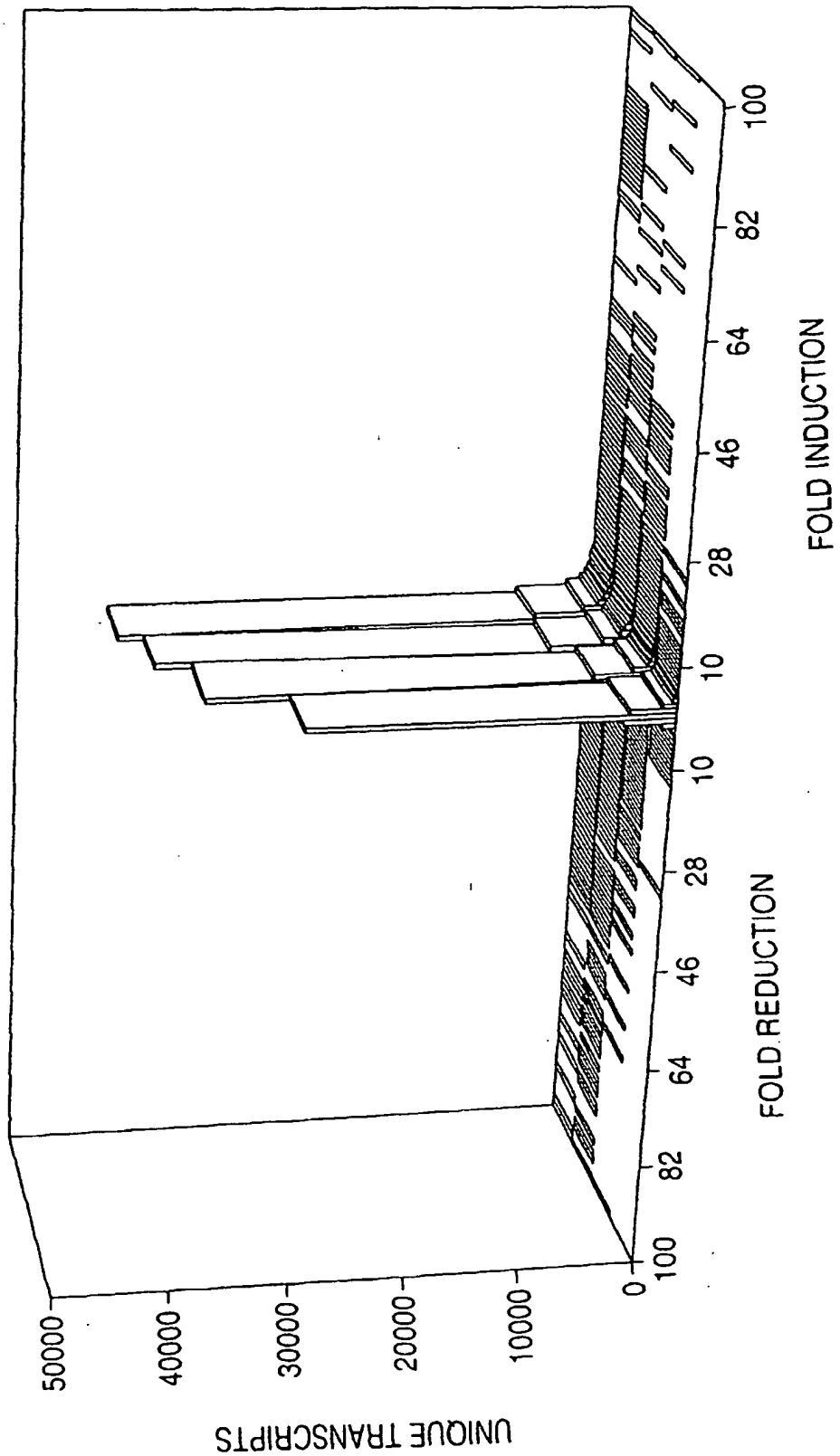
UNIQUE TRANSCRIPTS
50000
40000
30000
20000
10000
0
100
82
64
46
28
10
FOLD REDUCTION
10
28
46
64
82
100
FOLD INDUCTION

FIG. 3B

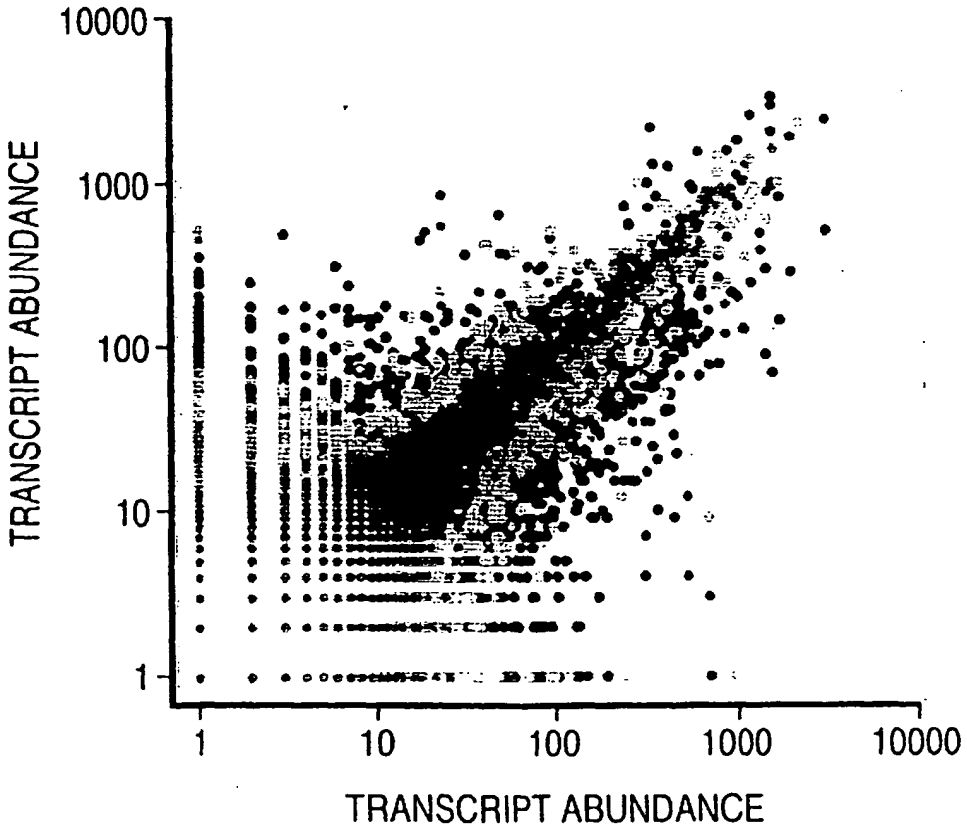
10000
1000
100
10
1
TRANSCRIPT ABUNDANCE
1
10
100
1000
10000
TRANSCRIPT ABUNDANCE

FIG. 3C

	COMPARISON	DESCRIPTION	TOTAL TRANSCRIPTS	UNIQUE TRANSCRIPTS	EXPRESSION CHANGE > = 10 FOLD (%)
☐	1	DLD1 CONDITION A vs DLD1 CONDITION B	0	42,673	43 (0.10)
☐	2	DLD1 vs HCT116	0	56,061	390 (0.70)[A]
▨	3	COLON CANCER vs NORMAL BRAIN	0	62,216	930 (1.49)[B]
▨	4	COLON CANCER vs HEMANGIOPERICYTOMA	0	72,239	1,047 (1.45)[C]

A DIFFERENCE BETWEEN EXPRESSION CHANGE OF COMPARISON 1 AND 2, $p < 0.0001$

B DIFFERENCE BETWEEN EXPRESSION CHANGE OF COMPARISON 2 AND 3, $p < 0.0001$

C DIFFERENCE BETWEEN EXPRESSION CHANGE OF COMPARISON 2 AND 4, $p < 0.0001$